

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)



Attorney Docket # 5394-3

Patent

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of

Gael MAHE et al.

Serial No.: 10/723,851

Filed: November 25, 2003

For: Method and System of Correcting Spectral
Deformations in the Voice, Introduced by
a Communication Network

Mail Stop
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

LETTER TRANSMITTING PRIORITY DOCUMENT

SIR:

In order to complete the claim to priority in the above-identified application under 35 U.S.C. §119, enclosed herewith is a certified copy of each foreign application on which the claim of priority is based: Application No. **02 15618**, filed on December 11, 2002, in France.

Respectfully submitted,
COHEN, PONTANI, LIEBERMAN & PAVANE

By Thomas Langer
Thomas Langer
Reg. No. 27,264
551 Fifth Avenue, Suite 1210
New York, New York 10176
(212) 687-2770

Dated: May 5, 2004



THIS PAGE BLANK (USPTO)



0215618



10/723,851

BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le **15 OCT. 2003**

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

SIEGE
26 bis, rue de Saint Petersburg
75800 PARIS cedex 08
Téléphone : 33 (0)1 53 04 53 04
Télécopie : 33 (0)1 53 04 45 23
www.inpi.fr

THIS PAGE BLANK (USPTO)



26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

BREVET D'INVENTION CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI

cerfa
N° 11354*01

REQUÊTE EN DÉLIVRANCE 1/2

Remplir impérativement la 2ème page.

Cet imprimé est à remplir lisiblement à l'encre noire

08 540 W / 190600

REMISE DES PIÈCES DATE 11 DEC 2002 LIEU 33 INPI BORDEAUX N° D'ENREGISTREMENT 0215618 NATIONAL ATTRIBUÉ PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI 11 DEC. 2002		1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE A. P. I. Conseil Lydie BORIN Rue Marx Dormoy 64000 PAU France	
Vos références pour ce dossier (facultatif) 0079-FT04453			
Confirmation d'un dépôt par télécopie <input type="checkbox"/> N° attribué par l'INPI à la télécopie			
2 NATURE DE LA DEMANDE		Cochez l'une des 4 cases suivantes	
Demande de brevet		<input checked="" type="checkbox"/>	
Demande de certificat d'utilité		<input type="checkbox"/>	
Demande divisionnaire		<input type="checkbox"/>	
<i>Demande de brevet initiale</i> <i>ou demande de certificat d'utilité initiale</i>		N° _____ Date ____/____/____ N° _____ Date ____/____/____	
Transformation d'une demande de brevet européen <i>Demande de brevet initiale</i>		<input type="checkbox"/> N° _____ Date ____/____/____	
3 TITRE DE L'INVENTION (200 caractères ou espaces maximum) PROCEDE ET SYSTEME DE CORRECTION MULTI-REFERENCES DES DEFORMATIONS SPECTRALES DE LA VOIX INTRODUITES PAR UN RESEAU DE COMMUNICATION.			
4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE		Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ Date ____/____/____ <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»	
5 DEMANDEUR		<input type="checkbox"/> S'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé «Suite»	
Nom ou dénomination sociale		FRANCE TELECOM	
Prénoms			
Forme juridique		Société Anonyme	
N° SIREN		
Code APE-NAF		
Adresse	Rue	6, Place d'Alleray	
	Code postal et ville	75015	PARIS
Pays		France	
Nationalité		Française	
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			



BREVET D'INVENTION CERTIFICAT D'UTILITÉ

REQUÊTE EN DÉLIVRANCE 2/2

REMISE DES PIÈCES DATE 11 DEC 2002 LIEU 33 INPI BORDEAUX N° D'ENREGISTREMENT 0215618 NATIONAL ATTRIBUÉ PAR L'INPI		Réservé à l'INPI	
Vos références pour ce dossier : <i>(facultatif)</i>		0079-FT04453	
6 MANDATAIRE			
Nom		BORIN	
Prénom		Lydie	
Cabinet ou Société		A.P.I. Conseil	
N° de pouvoir permanent et/ou de lien contractuel		94-0506	
Adresse	Rue	Marx Dormoy	
	Code postal et ville	64000	PAU
N° de téléphone <i>(facultatif)</i>		05 59 14 93 57	
N° de télécopie <i>(facultatif)</i>		05 59 14 43 76	
Adresse électronique <i>(facultatif)</i>		mail@api-conseil.com	
7 INVENTEUR (S)			
Les inventeurs sont les demandeurs		<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée	
8 RAPPORT DE RECHERCHE		Uniquement pour une demande de brevet (y compris division et transformation)	
Établissement immédiat ou établissement différé		<input checked="" type="checkbox"/> <input type="checkbox"/>	
Paiement échelonné de la redevance		Paiement en deux versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input type="checkbox"/> Non	
9 RÉDUCTION DU TAUX DES REDEVANCES		Uniquement pour les personnes physiques <input type="checkbox"/> Requête pour la première fois pour cette invention <i>(joindre un avis de non-imposition)</i> <input type="checkbox"/> Requête antérieurement à ce dépôt <i>(joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence)</i> :	
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes			
10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire) Lydie BORIN Mandataire N° 94-0506 A. P. I. Conseil		VISA DE LA PRÉFECTURE OU DE L'INPI 	

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.

L'invention concerne un procédé de correction multi références des déformations spectrales de la voix introduites par un réseau de communication. Elle concerne également un système pour la mise en œuvre du
5 procédé.

La présente invention a pour but l'amélioration de la qualité de la parole transmise sur les réseaux de communication, en offrant des moyens pour corriger les déformations spectrales du signal de parole,
10 déformations provoquées par différents maillons de la chaîne de transmission du réseau.

La description qui en est donnée dans la suite fait explicitement référence à la transmission de la parole sur lignes téléphoniques "classiques" (c'est-à-dire
15 filaires), mais elle s'applique également à tout type de réseau de communication (fixe, mobile ou autre) introduisant des déformations spectrales dans le signal, les paramètres pris comme référence pour qualifier le réseau devant être modifiés en fonction du
20 réseau.

On va rappeler dans ce qui suit les différentes déformations rencontrées dans le cas du Réseau Téléphonique Commuté (RTC).

1.1. Les dégradations du timbre de la voix sur le
25 Réseau RTC :

La figure 1 représente un schéma d'une liaison RTC. La parole émise par un locuteur est transmise par un terminal émetteur 10, transportée par la ligne d'abonné 20, subit une conversion 30 analogique-numérique (loi
30 A), est transmise par le réseau numérique 40, subit une conversion numérique (loi A)- analogique 50, est transmise par la liaison d'abonné 60, passe par le terminal récepteur 70 pour être enfin reçue par le destinataire.

Chaque locuteur est relié par une ligne analogique (paire torsadée) au central téléphonique le plus proche. Il s'agit d'une transmission analogique en bande de base référence 1 et 3 sur la figure 1. La liaison entre les centraux, emprunte un réseau
5 entièrement numérique 40. Le spectre de la voix est affecté par deux types de distorsions lors de la transmission analogique du signal en bande de base.

Le premier type de distorsion est le filtrage passe-bande des terminaux et des points d'accès à la
10 partie numérique du réseau. Les caractéristiques typiques de ce filtrage sont décrites par l'UIT-T sous le nom de "système de référence intermédiaire" (SRI) [UIT-T, Recommandation P.48, 1988]. Ces
15 caractéristiques fréquentielles, issues de mesures réalisées dans les années 70, tendent cependant à devenir obsolètes. C'est pourquoi l'UIT-T préconise depuis 1996 d'utiliser un SRI "modifié" [UIT-T, Recommandation P.830, 1996], dont la caractéristique
20 nominale est représentée sur la figure 2 pour la partie émission, et sur la figure 3 pour la partie réception. Entre 200 et 3400 Hz, la tolérance est de $\pm 2,5$ dB ; en dessous de 200 Hz, la décroissance de la
25 caractéristique du système global doit être d'au moins 15 dB par octave. On appelle respectivement, selon la terminologie de l'UIT-T, "système émetteur" et "système récepteur" les parties émission et réception du SRI.

La seconde distorsion affectant le spectre de la voix est l'atténuation des lignes d'abonné. Dans un
30 modèle simple de la ligne analogique locale [donné dans une Note Technique CNET NT/LAA/ELR/289 par Cadoret, 1983], on considère que celle-ci introduit un affaiblissement du signal dont la valeur en dB dépend de sa longueur et est proportionnelle à la racine

carrière de la fréquence. L'affaiblissement est de 3 dB à 800 Hz pour une ligne moyenne (environ 2 km), de 9,5 dB à 800 Hz pour les lignes les plus longues (jusqu'à 10 km). Selon ce modèle, l'affaiblissement d'une ligne, représenté sur la figure 4, a pour expression :

$$A_{dB}(f) = A_{dB}(800\text{Hz}) \cdot \sqrt{\frac{f}{800}} \quad (0.1)$$

A ces distorsions s'ajoute le filtrage anti-repliement du codeur MIC (réf 30). Celui-ci est typiquement un filtre passe-bande 200-3400 Hz avec une réponse presque plate sur la bande passante et une forte atténuation en dehors de la bande, selon le gabarit de la figure 5 par exemple [National Semiconductor, août 1994 : Documentation technique « TP3054 ,TP3057 »].

Au final, la voix subit une distorsion spectrale telle que représentée sur la figure 6 pour les différentes combinaisons de trois types de ligne analogique en émission et en réception (soit 6 distorsions), sous l'hypothèse d'équipements respectant la caractéristique nominale du SRI modifié. La voix apparaît ainsi étouffée si une des lignes analogiques est longue et souffre dans tous les cas d'un manque de "présence" dû à l'affaiblissement des composantes basse fréquence.

1.2. Les dégradations du timbre de la voix sur le Réseau Numérique à Intégration de Services (RNIS) et le réseau mobile GSM.

Dans le RNIS et le réseau GSM, le signal est numérisé dès le terminal. Les seules parties analogiques sont les transducteurs en émission et en réception associés à leurs chaînes d'amplification et de conditionnement respectives. L'UIT-T a défini des

gabarits d'efficacité en fréquence à l'émission représentés sur la figure 7, et à la réception représenté sur la figure 8, valables à la fois pour les téléphones numériques filaires [UIT-T, Recommandation P.310, mai 2000] et les terminaux numériques mobiles ou sans fil [UIT-T, Recommandation P.313, septembre 1999].

Par ailleurs, pour les réseaux GSM, il est reconnu que le codage et le décodage modifient légèrement l'enveloppe spectrale du signal. Cette altération est représentée sur la figure 9 pour un bruit rose codé puis décodé en mode EFR (Enhanced Full Rate).

L'effet de ces filtrages sur le timbre est principalement un affaiblissement des composantes basse fréquence, moins marqué cependant que dans le cas du RTC.

L'invention concerne la correction de ces distorsions spectrales par un traitement centralisé, c'est-à-dire un dispositif installé dans la partie numérique du réseau, comme indiqué sur la figure 10 pour le RTC.

L'objectif d'une correction du timbre de la voix est que le timbre de la voix en réception soit le plus proche possible de celui de la voix émise par le locuteur, que l'on appellera voix originale.

2. État de la technique

Une compensation des distorsions spectrales introduites dans le signal de parole par les divers éléments de la liaison téléphonique est permise à ce jour par des dispositifs à base d'égalisation. Celle-ci peut être fixe ou s'adapter en fonction des conditions de transmission.

2.1 L'égalisation fixe

Des dispositifs d'égalisation centralisée ont été proposés dans les brevets US 5333195 [Duane O. Bowker] et US 5471527 [Helena S. Ho]. Ces égaliseurs sont des filtres fixes qui restaurent le niveau des basses fréquences atténuées par l'émetteur. Bowker propose par exemple un gain de 10 à 15 dB sur la bande 100-300 Hz. Ces méthodes présentent deux inconvénients :

* L'égaliseur ne compense que le filtrage de l'émetteur, de sorte qu'à la réception, les composantes basse-fréquence restent fortement affaiblies par le filtrage SRI de réception.

* Cette égalisation fixe compense des conditions de transmission (ligne et système d'émission) moyennes. Si les conditions réelles sont trop différentes (par exemple si les lignes analogiques sont longues) le dispositif ne corrige pas suffisamment le timbre, voire l'altère plus que la liaison sans égalisation.

2.2 L'égalisation adaptative.

Le dispositif décrit dans le brevet US 5915235 [Andrew P De Jaco] vise à corriger la réponse fréquentielle non idéale d'un transducteur de téléphone mobile. L'égaliseur est décrit comme étant placé entre le convertisseur analogique-numérique et le codeur CELP, mais peut être aussi bien dans le terminal que dans le réseau. Le principe de l'égalisation est de rapprocher le spectre du signal reçu d'un spectre idéal. Deux méthodes sont proposées.

La première méthode (illustrée par la figure 4 du brevet précité de De Jaco), consiste à calculer des coefficients d'autocorrélation à long terme RLT :

$$R_{LT}(n,i) = \alpha R_{LT}(n-1,i) + (1-\alpha)R(n,i), \quad (0.2)$$

avec $R_{LT}(n,i)$ $i^{\text{ième}}$ coefficient d'autocorrélation à long terme à la $n^{\text{ième}}$ trame, $R(n,i)$ $i^{\text{ième}}$ coefficient d'autocorrélation spécifique à la $n^{\text{ième}}$ trame, et α

constante de lissage fixée par exemple à 0,995. De ces coefficients sont dérivés les coefficients LPC à long terme, qui sont les coefficients d'un filtre blanchisseur. À la sortie de ce filtre, le signal est
5 filtré par un filtre fixe qui lui imprime les caractéristiques spectrales à long terme idéales, ie celles qu'il aurait à la sortie d'un transducteur ayant la réponse fréquentielle idéale. Ces deux filtres sont complétés par un gain multiplicatif égal au rapport
10 entre les énergies à long terme de l'entrée du blanchisseur et de la sortie du deuxième filtre.

La deuxième méthode, illustrée par la figure 5 du brevet précité de De Jaco, consiste à diviser le signal en sous-bandes, et, pour chaque sous-bande, appliquer
15 un gain multiplicatif de manière à atteindre une énergie cible, ce gain étant défini comme le rapport entre l'énergie cible de la sous-bande et l'énergie à long terme (obtenue par un lissage de l'énergie instantanée) du signal dans cette sous-bande.

20 Ces deux méthodes présentent l'inconvénient de ne corriger que la réponse non idéale du système d'émission, et pas celle du système de réception.

Le dispositif du brevet US 5905969 [Chafik Mokbel] a pour objet de compenser le filtrage du système
25 d'émission et de la ligne d'abonné pour améliorer la reconnaissance centralisée de la parole et/ou la qualité de la parole transmise. Comme le présente la figure 3a de Mokbel, le spectre du signal est divisé en 24 sous-bandes, et chaque énergie de sous-bande est
30 multipliée par un gain adaptatif. L'adaptation du gain est réalisée selon l'algorithme du gradient stochastique, par minimisation de l'erreur quadratique, l'erreur étant définie comme la différence entre l'énergie de sous-bande et une énergie de référence

définie pour chaque sous-bande. L'énergie de référence est modulée à chaque trame par l'énergie de la trame courante, de manière à respecter les variations naturelles de niveau à court terme du signal de parole. La convergence de l'algorithme permet d'obtenir en

sortie les 24 signaux de sous-bande égalisés. Si l'application visée est l'amélioration de la qualité vocale, le signal de parole égalisé est obtenu par transformée de Fourier inverse des énergies de sous-bande égalisées.

Le brevet de Mokbel, ne mentionne pas de résultats en termes d'amélioration de la qualité vocale, et reconnaît que la méthode est sous-optimale, en ce qu'elle réalise une convolution circulaire. Par ailleurs, il est douteux qu'un signal de parole puisse être reconstruit correctement par transformée de Fourier inverse d'énergies de bandes distribuées selon l'échelle MEL. Enfin, le dispositif décrit ne corrige pas le filtrage du système de réception et de la ligne analogique de réception.

La compensation de l'effet de ligne est réalisée dans la méthode « Mokbel », de soustraction cepstrale, dans le but d'améliorer la robustesse de la reconnaissance de la parole. On montre que le cepstre du canal de transmission peut être estimé par le cepstre moyen du signal reçu, celui-ci étant préalablement blanchi par un filtre de préaccentuation. Cette méthode permet une nette amélioration des performances des systèmes de reconnaissance, mais est considérée comme une méthode "off-line", 2 à 4 s étant nécessaires pour estimer le cepstre moyen.

2.3 Un autre état de la technique combine une pré-égalisation fixe et une égalisation adaptée et a fait l'objet du dépôt d'une demande de brevet FR 2822999 par

le Déposant. Le dispositif décrit vise à corriger le timbre de la parole en combinant deux filtres.

Un filtre fixe, appelé pré-égaliseur, compense les distorsions d'une liaison téléphonique moyenne, définie comme étant constituée de deux lignes d'abonné moyennes et de systèmes d'émission et de réception respectant les réponses fréquentielles nominales définies dans [UIT-T, Recommandation P.48, App. I, 1988]. Sa réponse fréquentielle, sur la bande $[F_c - 3150 \text{ Hz}]$, est l'inverse de la réponse globale de la partie analogique de cette liaison moyenne, F_c étant la fréquence basse limite d'égalisation.

Cette pré-égalisation est complétée par un égaliseur adapté, qui adapte la correction de manière plus précise aux conditions réelles de transmission. La réponse fréquentielle de l'égaliseur adapté est donnée par :

$$|EQ(f)| = \frac{1}{|S_{RX}(f) L_{RX}(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \quad (0.3)$$

avec L_{RX} la réponse fréquentielle de la ligne de réception, S_{RX} la réponse fréquentielle du système de réception et $\gamma_x(f)$ le spectre à long terme de la sortie x du pré-égaliseur.

Le spectre à long terme est défini comme la moyenne temporelle des spectres à court terme des trames successives de signal ; $\gamma_{ref}(f)$, appelé spectre de référence, est le spectre moyen de la parole défini par l'UIT [UIT-T/P.50/App. I, 1998], pris comme approximation du spectre à long terme original du locuteur. Du fait de cette approximation, la réponse fréquentielle de l'égaliseur adapté est très irrégulière et seule sa forme générale est pertinente. C'est pourquoi elle doit être lissée. L'égaliseur adapté étant réalisé sous la forme d'un filtre temporel

RIF, ce lissage dans le domaine fréquentiel est obtenu par un étroit fenêtrage (symétrique) de la réponse impulsionnelle.

5 Cette méthode permet de restaurer un timbre proche de celui du signal original sur la bande d'égalisation [Fc-3150 Hz], mais :

10 - pour certains locuteurs, l'approximation de leur spectre à long terme original par le spectre de référence est très grossière, de sorte que l'égaliseur introduit une distorsion perceptible ;

15 - le fort lissage de la réponse fréquentielle de l'égaliseur, rendu nécessaire par l'erreur d'approximation, interdit de corriger des distorsions spectrales fines.

15

20 L'invention a pour but de remédier aux inconvénients de l'état de la technique. Elle a pour objet un procédé et un système pour améliorer la correction du timbre en réduisant l'erreur d'approximation du spectre à long terme original des locuteurs.

25 A cet effet, il est proposé de classer les locuteurs selon leur spectre à long terme et d'approcher celui-ci non plus par un spectre de référence unique mais par un spectre de référence par classe. Le procédé proposé permet de réaliser un traitement d'égalisation apte à déterminer la classe du locuteur et à égaliser suivant le spectre de référence de la classe. Cette réduction de l'erreur
30 d'approximation permet de lisser moins fortement la réponse fréquentielle de l'égaliseur adapté, le rendant apte à corriger des distorsions spectrales plus fines.

La présente invention a plus particulièrement pour objet un procédé de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant une opération d'égalisation sur une bande de fréquence $[F1-F2]$, adaptée à la distorsion réelle de la chaîne de transmission, cette opération étant réalisée au moyen d'un filtre numérique ayant une réponse fréquentielle fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal voix des locuteurs, principalement caractérisé en ce qu'il comprend :

* Préalablement à l'opération d'égalisation du signal voix d'un locuteur en communication :

15 - La constitution de classes de locuteurs avec une référence de voix par classe,

* Puis, pour un locuteur donné en communication :

20 - Le classement de ce locuteur c'est-à-dire son affectation à une classe à partir de critères de classement prédéfinis pour lui faire correspondre une référence de voix qui est la plus proche de la sienne,

25 - L'égalisation du signal numérisé de la voix du locuteur opérée avec comme spectre de référence, la référence de voix de la classe à laquelle ledit locuteur a été affecté.

Selon une autre caractéristique, la constitution de classes de locuteurs comprend :

30 - Le choix d'un corpus de N locuteurs enregistrés dans des conditions non dégradées et la détermination de leur spectre de fréquence à long terme,

- La classification des locuteurs du corpus selon leur cepstre partiel c'est à dire le cepstre calculé à partir du spectre à long terme restreint à la bande d'égalisation $[F1-F2]$ et en appliquant un critère de

classification prédéfini sur ces cepstres pour obtenir K classes,

- Le calcul du spectre de référence associé à chaque classe de manière à obtenir une référence de voix correspondant à chacune des classes.

Selon une autre caractéristique, le spectre de référence sur la bande de fréquences d'égalisation [F1-F2], associé à chaque classe, est calculé par transformée de Fourier du centre de la classe défini par son cepstre partiel.

Selon une autre caractéristique, la classification d'un locuteur comprend :

- L'utilisation du pitch moyen du signal voix et du cepstre partiel de ce signal comme paramètres de classement,

- L'application d'une fonction discriminante à ces paramètres pour classer ledit locuteur.

Selon l'invention le procédé comprend en outre une étape de pré-égalisation du signal numérique par un filtre fixe ayant une réponse fréquentielle dans la bande de fréquence [F1-F2], correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison téléphonique.

Selon une autre caractéristique, l'égalisation du signal numérisé de la voix d'un locuteur comprend :

- la détection d'une activité vocale sur la ligne pour déclencher un enchaînement de traitements comprenant le calcul du spectre à long terme, la classification du locuteur, le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] et le calcul des coefficients du filtre numérique différenciés selon la classe du locuteur, à partir de ce module,

- la commande du filtre avec les coefficients obtenus,

- le filtrage du signal sortant du pré-égaliseur par ledit filtre.

5 Selon une autre caractéristique, le calcul du module [EQ] de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] est réalisé par la mise en œuvre de la relation suivante :

10

$$|EQ(f)| = \frac{1}{|S_{RX}(f) \cdot L_{RX}(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \quad (0.3)$$

dans laquelle $\gamma_{ref}(f)$ est le spectre de référence de la classe à laquelle appartient ledit locuteur, et dans laquelle L_{RX} est la réponse fréquentielle de la ligne de réception, S_{RX} la réponse fréquentielle du système de réception et $\gamma_x(f)$ le spectre à long terme du signal x d'entrée du filtre.

15

20 Selon une variante, le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] est réalisé par la mise en œuvre de la relation suivante :

25

$$C_{eq}^p = C_{ref}^p - C_x^p - C_{s_{rx}}^p - C_{l_{rx}}^p, \quad (0.13)$$

dans laquelle C_{eq}^p , C_x^p , $C_{s_{rx}}^p$ et $C_{l_{rx}}^p$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C_{ref}^p étant le cepstre partiel de référence, centre de la classe du locuteur. Le module [EQ] restreint à la bande F1-F2 est alors calculé par Transformée de Fourier discrète de C_{eq}^p .

30

L'invention a également pour objet un système de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant des moyens d'égalisation adaptée dans une bande de fréquence [F1-F2] qui comportent un filtre numérique dont la réponse fréquentielle est fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme d'un signal voix, principalement caractérisé en ce que ces moyens comprennent en outre :

- des moyens de traitements du signal pour le calcul des coefficients du filtre numérique munis :

- d'un bloc de traitement du signal pour calculer le module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] selon la relation suivante :

$$|EQ(f)| = \frac{1}{|S_{RX}(f) L_{RX}(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \quad (0.3)$$

dans laquelle $\gamma_{ref}(f)$ est le spectre de référence, qui peut être différent d'un locuteur à un autre et qui correspond à une référence de classe prédéterminée à laquelle appartient ledit locuteur, et dans laquelle L_{RX} est la réponse fréquentielle de la ligne de réception, S_{RX} la réponse fréquentielle du système de réception et $\gamma_x(f)$ le spectre à long terme du signal x d'entrée du filtre;

- d'un deuxième bloc de traitements pour le calcul de la réponse impulsionnelle à partir du module de réponse fréquentielle ainsi calculé,

afin de déterminer les coefficients du filtre différenciés selon la classe du locuteur.

5 Selon une autre caractéristique, le premier bloc de traitement comprend des moyens de calcul du cepstre partiel du filtre égaliseur selon la relation :

$$C_{eq}^p = C_{ref}^p - C_x^p - C_{s_{rx}}^p - C_{l_{rx}}^p, \quad (0.13)$$

10 dans laquelle C_{eq}^p , C_x^p , $C_{s_{rx}}^p$ et $C_{l_{rx}}^p$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C_{ref}^p étant le cepstre partiel de référence, centre de la classe du locuteur, le module de [EQ] restreint à la bande F1-F2 est alors calculé par Transformée de Fourier Discrète de C_{eq}^p .

15 Selon une autre caractéristique, le premier bloc de traitements comprend un sous-ensemble pour calculer les coefficients du cepstre partiel d'un locuteur en communication et un deuxième sous-ensemble pour opérer le classement de ce locuteur, ce deuxième sous-ensemble
20 comprenant un bloc de calcul du pitch F_0 , un bloc d'estimation du pitch moyen à partir du pitch calculé F_0 , et un bloc de classement appliquant une fonction discriminante sur le vecteur x ayant pour composantes le pitch moyen et les coefficients du cepstre partiel
25 pour classer ledit locuteur.

Selon l'invention, le système comprend en outre un pré-égaliseur, le signal égalisé à partir de spectres de référence différenciés selon la classe du locuteur étant le signal x de sortie du pré-égaliseur.

30

D'autres particularités et avantages de l'invention apparaîtront clairement dans la description suivante, qui est donnée à titre d'exemple illustratif

et non limitatif et qui est faite en regard des figures annexées qui représentent :

- La figure 1, une liaison téléphonique schématisée pour un réseau téléphonique commuté (RTC),

5 - La figure 2, la courbe de réponse fréquentielle en émission du système de référence intermédiaire SRI modifié,

10 - La figure 3, la courbe de réponse fréquentielle en réception du système de référence intermédiaire SRI modifié,

- La figure 4, la réponse en fréquence des lignes d'abonnés selon leur longueur,

- La figure 5, le gabarit du filtre anti-repliement du codeur MIC,

15 - La figure 6, les distorsions spectrales subies par la parole sur le réseau téléphonique commuté avec SRI moyen et différentes combinaisons de lignes analogiques,

- La figure 7, le gabarit à l'émission pour les terminaux numériques,

20 - La figure 8, le gabarit à la réception pour les terminaux numériques,

- La figure 9, la distorsion spectrale introduite par le codage-décodage GSM en mode EFR (Enhanced Full Rate)

25 - La figure 10, le schéma d'un réseau de communication avec un système de correction des distorsions de la parole,

- La figure 11, les étapes de calcul du cepstre partiel,

30 - La figure 12, la classification des cepstres partiels selon le critère de la variance,

- Les figures 13a et 13b, les spectres à long terme correspondant aux centres des classes de locuteurs respectivement pour hommes et femmes,

5 - La figure 14, les caractéristiques fréquentielles des filtrages appliqués au corpus pour définir le corpus d'apprentissage,

- La figure 15, la réponse fréquentielle du pré-égaliseur pour différentes fréquences F_c ,

10 - La figure 16, le schéma de réalisation du système de correction par égalisation différenciée par classe de locuteurs,

- La figure 17, une variante d'exécution du système selon la figure 16.

15 Dans toute la suite les mêmes références portées sur les dessins correspondent aux mêmes éléments.

La description qui va suivre va d'abord présenter l'étape préalable de classification d'un corpus de locuteurs selon leur spectre à long terme. Cette étape
20 permet de définir K classes et une référence par classe.

Un enchaînement de traitements permet de traiter le signal de parole (dès la détection d'une activité vocale par le système) de chaque locuteur pour d'une
25 part classer les locuteurs c'est à dire les affecter à une classe selon des critères prédéterminés et pour d'autre part corriger la voix en utilisant la référence de la classe du locuteur.

30 Etape préalable de classification des locuteurs.

* Choix du corpus de définition des classes.

Le spectre de référence étant une approximation du spectre à long terme original des locuteurs, la définition des classes de locuteurs et de leurs

spectres de référence respectifs nécessite de disposer d'un corpus de locuteurs enregistrés dans des conditions non dégradées. En particulier, le spectre à long terme d'un locuteur mesuré sur cet enregistrement doit pouvoir être considéré comme son spectre original, i.e. celui de sa voix à l'extrémité d'émission d'une liaison téléphonique.

Définition de l'individu : le cepstre partiel.

Le traitement proposé permet de disposer, dans chaque classe, d'un spectre de référence le plus proche possible du spectre à long terme de chaque membre de la classe. Cependant, seule la partie du spectre comprise dans la bande d'égalisation F1-F2 est prise en compte dans le traitement d'égalisation adaptée. Les classes sont donc constituées selon le spectre à long terme restreint à cette bande.

D'autre part, la comparaison entre deux spectres est effectuée à un faible niveau de résolution spectrale, de manière à ne refléter que l'enveloppe spectrale. C'est pourquoi, on se place de préférence, dans l'espace des premiers coefficients cepstraux d'ordre supérieur à 0 (le coefficient d'ordre 0 représentant l'énergie), le choix du nombre de coefficients dépendant de la résolution spectrale souhaitée.

On détermine donc dans le traitement, le "cepstre partiel à long terme", que l'on note C_p , comme la représentation cepstrale du spectre à long terme restreint à une bande de fréquence. Si l'on note k_1 et k_2 les indices de fréquence correspondant respectivement aux fréquences F1 et F2, et γ le spectre à long terme de la parole, le cepstre partiel est défini par la relation :

$$C^p = \text{TFD}^{-1} \left(10 \log \left(\gamma(k_1 \dots k_2) \circ \gamma(k_2 - 1 \dots k_1 + 1) \right) \right) \quad (0.4)$$

où \circ désigne l'opération de concaténation.

La (TFD) Transformée de Fourier Discrète Inverse est calculée par exemple par IFFT après interpolation des échantillons du spectre tronqué de manière à
5 atteindre un nombre d'échantillons puissance de 2. Par exemple, en choisissant la bande d'égalisation 187-3187 Hz, correspondant aux indices fréquentiels 5 à 101 pour une représentation du spectre (symétrisé) sur 256 points (de 0 à 255), l'interpolation se fait simplement
10 en intercalant une raie fréquentielle (interpolée linéairement) toutes les trois raies dans le spectre restreint à 187-3187 Hz.

Les étapes du calcul du cepstre partiel sont représentées sur la figure 11.

15 Pour que les coefficients cepstraux reflètent l'enveloppe spectrale mais pas l'influence de la structure harmonique du spectre de la parole sur les spectres à long terme, on ne conserve pas les coefficients d'ordre élevé. Les locuteurs à classer
20 sont donc représentés par les coefficients d'ordres 1 à L de leur cepstre partiel à long terme, L valant typiquement 20.

* La Classification.

25 Les classes sont constituées par exemple de manière non-supervisée, selon une classification hiérarchique ascendante.

Celle-ci consiste à créer, à partir de N individus disjoints, une hiérarchie de partitions selon le processus suivant : à chaque étape, on agrège les deux
30 éléments les plus proches, un élément étant soit un individu non agrégé, soit un agrégat d'individus constitué lors d'une précédente étape. La proximité entre deux éléments est déterminée par une mesure de dissimilarité que l'on appelle distance. Le processus

se poursuit jusqu'à l'agrégation de toute la population. La hiérarchie de partitions ainsi créée peut se représenter sous la forme d'un arbre comme celui de la figure 12, contenant N-1 partitions
 5 imbriquées. Chaque coupure de l'arbre fournit une partition, d'autant plus fine que l'on coupe bas.

Dans ce type de classification, on choisit comme mesure de distance entre deux éléments, la variation d'inertie intra-classes résultant de leur agrégation.
 10 Une partition est en effet d'autant meilleure que les classes créées sont homogènes, c'est-à-dire que l'inertie intra-classes est faible. Dans le cas d'un nuage de points x_i de masses respectives m_i , répartis en classes q de centres de gravité respectifs g_q ,
 15 l'inertie intra-classes est définie par :

$$I_{\text{intra}} = \sum_q \sum_{i \in q} m_i \|x_i - g_q\|^2. \quad (0.5)$$

L'inertie intra-classes, nulle à l'étape initiale de l'algorithme de calcul, s'accroît inévitablement à chaque agrégation.

20 On utilise de préférence le principe connu de l'agrégation selon la variance. Selon ce principe on recherche, à chaque étape de l'algorithme mis en oeuvre, les deux éléments dont l'agrégation produit l'augmentation d'inertie intra-classes la plus faible.

25 La partition ainsi obtenue est améliorée par une procédure d'agrégation autour des centres mobiles, qui permet de réduire la variance intra-classes.

Le spectre de référence, sur la bande F1-F2, associé à chaque classe est calculé par transformée de
 30 Fourier du centre de la classe.

* Exemple de classification

Le traitement décrit ci-dessus est appliqué à un corpus de 63 locuteurs. L'arbre de classification du

corpus est représenté sur la figure 12. Dans cette représentation, la hauteur d'un segment horizontal agrégeant deux éléments est choisie proportionnelle à leur distance, ce qui permet de visualiser la proximité des éléments regroupés dans une même classe. Cette représentation facilite le choix du niveau de coupure de l'arbre, et donc des classes retenues. La coupure doit être faite au-dessus des agrégations de niveau faible, qui regroupent des individus proches, et en dessous des agrégations de niveau élevé, qui associent des groupes d'individus bien distincts.

De cette manière, on obtient de façon nette quatre classes ($K=4$). Ces classes sont assez homogènes du point de vue du sexe des locuteurs, et une coupure de l'arbre en deux classes fait apparaître à peu près une classe hommes et une classe femmes.

La consolidation de cette partition par une procédure d'agrégation autour des centres mobiles aboutit à quatre classes de cardinaux 11, 18, 18 et 16, plus homogènes que précédemment du point de vue du sexe : seuls un homme et deux femmes sont affectés à des classes ne correspondant pas à leur sexe.

Les spectres restreints à la bande 187-3187 Hz correspondant aux centres de ces classes sont représentés sur les figures 13a et 13b pour les classes hommes et femmes ainsi que pour leurs sous-classes respectives. Ces spectres, résultats de la classification, sont utilisés comme référence multiple par l'égaliseur adapté.

* Utilisation de critères de classement des locuteurs.

Les classes de locuteurs étant définies, le traitement prévoit l'utilisation de paramètres et des

critères pour affecter un locuteur à l'une ou l'autre des classes.

Cette affectation n'est pas réalisée simplement selon la proximité du cepstre partiel avec un des centres de classes, puisque ce cepstre est dévié par la partie de la liaison téléphonique en amont de l'égaliseur.

Il est proposé avantageusement d'utiliser des critères de classement robustes à cette déviation. Cette robustesse est assurée à la fois par le choix des paramètres de classement et par celui du corpus d'apprentissage des critères de classement.

* De préférence on utilise les paramètres de classement : pitch moyen et cepstre partiel.

Les classes précédemment définies sont homogènes du point de vue du sexe. Le pitch moyen étant à la fois assez discriminant pour un classement homme / femme et insensible aux distorsions spectrales induites par une liaison téléphonique, il est donc utilisé comme paramètre de classement, conjointement avec le cepstre partiel.

* Choix du corpus d'apprentissage des critères de classement.

On applique à ces paramètres une technique de discrimination par exemple la technique usuelle de l'analyse linéaire discriminante.

D'autres techniques connues peuvent être utilisées telles qu'une technique non linéaire utilisant un réseau de neurones.

Si l'on dispose de N individus décrits par des vecteurs de dimension p et répartis a priori en K classes, l'analyse linéaire discriminante consiste :

- dans un premier temps, à chercher les $K-1$ fonctions linéaires indépendantes qui séparent au mieux

les K classes. Il s'agit de déterminer quelles sont les combinaisons linéaires des p composantes des vecteurs qui minimisent la variance intra-classes et maximisent la variance interclasses.

- 5 - dans un deuxième temps, à déterminer la classe d'un nouvel individu par application des fonctions linéaires discriminantes au vecteur le représentant.

10 Dans le cas présent, les vecteurs représentatifs des individus ont pour composantes le pitch et les coefficients 1 à L (typiquement, $L = 20$) du cepstre partiel. La robustesse des fonctions discriminantes à la déviation des coefficients cepstraux est assurée à la fois par la présence du pitch dans les paramètres et par le choix du corpus d'apprentissage. Celui-ci est
15 composé d'individus dont la voix originale a subi une grande diversité de filtrages représentatifs des distorsions occasionnées par les liaisons téléphoniques.

20 Plus précisément, à partir d'un corpus de voix originales (non dégradées) de N locuteurs, on définit un corpus de N vecteurs de composantes $[\bar{F}_0; C^p(1); \dots; C^p(L)]$, avec \bar{F}_0 le pitch moyen et C^p le cepstre partiel. La construction du corpus d'apprentissage desdites fonctions consiste à définir un ensemble de M biais
25 cepstraux qui s'ajouteront chacun à chaque cepstre partiel représentatif d'un locuteur du corpus original, ce qui permet d'obtenir un nouveau corpus de NM individus.

30 Ces biais dans le domaine du cepstre partiel correspondent à une large gamme de distorsions spectrales sur la bande F1-F2, proches de celles pouvant résulter de la liaison téléphonique.

A titre d'exemple, on propose l'ensemble de réponses fréquentielles représentées sur la figure 14

pour la bande 187-3187 Hz : chaque réponse fréquentielle correspond à un chemin de gauche à droite dans le treillis. L'amplitude de leurs variations sur cette bande n'excède pas 20 dB, à l'instar des caractéristiques extrémales des systèmes d'émission et lignes.

A partir de ces 81 caractéristiques fréquentielles sont calculés les 81 biais correspondants dans le domaine du cepstre partiel, selon le traitement décrit pour la mise en œuvre de la relation (0.4). Par addition de ces biais au corpus de 63 locuteurs précédemment utilisé, on obtient un corpus d'apprentissage comptant 5103 individus représentatifs de diverses conditions (locuteur, filtrage de la liaison).

Dans le cas de classement par analyse linéaire discriminante :

* Application des critères de classement.

Soit $(a^k)_{1 \leq k \leq K-1}$ la famille de fonctions linéaires discriminantes définies à partir du corpus d'apprentissage. Un locuteur représenté par le vecteur $x = [\bar{F}_0; C^p(1); \dots; C^p(L)]$ est affecté à la classe q si la probabilité conditionnelle de q sachant $a(x)$, notée $P(q|a(x))$, est maximale, $a(x)$ désignant le vecteur de composantes $(a^k(x))_{1 \leq k \leq K-1}$. Selon le théorème de Bayes,

$$P(q|a(x)) = \frac{P(a(x)|q)P(q)}{P(a(x))} \quad (0.6)$$

Par conséquent, $P(q|a(x))$ est proportionnelle à $P(a(x)|q)P(q)$. Dans le sous-espace engendré par les $K-1$ fonctions discriminantes, sous l'hypothèse d'une distribution multi-gaussienne des individus dans chaque classe, la densité de probabilité de $a(x)$ à l'intérieur de la classe q a pour expression :

$$f_q(x) = \frac{1}{(2\pi)^{\frac{K-1}{2}} \sqrt{|S_q|}} \exp\left(-\frac{1}{2}(a(x)-a(\bar{x}^q))' S_q^{-1} (a(x)-a(\bar{x}^q))\right), \quad (0.7)$$

où \bar{x}^q est le centre de la classe q , $|S_q|$ désigne le déterminant de la matrice S_q , et S_q est la matrice des covariances de \underline{a} à l'intérieur de la classe q , d'élément générique σ_{jk}^q que l'on peut estimer par :

$$\sigma_{jk}^q = \frac{1}{N_q} \sum_{i=1}^{N_q} (a^j(x^i) - a^j(\bar{x}^q))(a^k(x^i) - a^k(\bar{x}^q)). \quad (0.8)$$

L'individu x sera affecté à la classe q qui maximise $f_q(x)P(q)$, ce qui revient à minimiser sur q la fonction $s_q(x)$ appelée score discriminant :

$$s_q(x) = (a(x) - a(\bar{x}^q))' S_q^{-1} (a(x) - a(\bar{x}^q)) + \log(|S_q|) - 2\log(P(q)), \quad (0.9)$$

Le procédé de correction proposé est mis en œuvre par le système de correction (égaliseur) implanté dans le réseau numérique 40 comme l'illustre la figure 10.

La figure 16 illustre le système de correction apte à mettre en œuvre le procédé. La figure 17 illustre ce système selon une variante de réalisation comme cela va être détaillé dans la suite. Ces variantes portent sur le mode de calcul du module de la réponse fréquentielle de l'égaliseur adapté restreinte à la bande F1-F2.

Le pré-égaliseur 200 est un filtre fixe, dont la réponse fréquentielle, sur la bande F1-F2, est l'inverse de la réponse globale de la partie analogique d'une liaison moyenne telle que définie précédemment [UIT-T/P.830, 1996].

La raideur de la réponse fréquentielle de ce filtre implique une réponse impulsionnelle longue ; c'est pourquoi, de manière à limiter le retard introduit par le traitement, le pré-égaliseur est réalisé typiquement sous forme d'un filtre RII, d'ordre 20 par exemple.

La figure 15 représente les réponses fréquentielles typiques du pré-égaliseur pour trois valeurs de F_1 . La dispersion des retards de groupe est inférieure à 2 ms, de sorte que la distorsion de phase résultante n'est pas perceptible.

5 La chaîne de traitement 400 qui suit permet le classement du locuteur et l'égalisation adaptée différenciée. Cette chaîne comprend deux blocs de traitements 400A et 400B. Le bloc 400A permet de
10 calculer le module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation : $EQ_{dB} [F_1-F_2]$.

Le deuxième bloc 400B permet de calculer la réponse impulsionnelle du filtre égaliseur afin d'obtenir les
15 coefficients $eq(n)$ du filtre différenciés selon la classe du locuteur.

Un détecteur de trames d'activité vocale 401 permet de déclencher les différents traitements.

Le bloc de traitement 410 permet le classement du
20 locuteur.

Le bloc de traitement 420 permet de calculer le spectre à long terme suivi du calcul du cepstre partiel de ce locuteur.

La sortie de ces deux blocs est appliquée sur
25 l'opérateur 428a ou 428b. La sortie de cet opérateur fournit le module de la réponse fréquentielle de l'égaliseur adapté en dB, restreinte à la bande d'égalisation F_1-F_2 via le bloc 429 pour 428a, via le bloc 440 pour 428b.

30 Les blocs de traitement 430 à 435 permettent de calculer les coefficients $eq(n)$ du filtre.

La sortie $x(n)$ du pré-égaliseur est analysée par trames successives d'une durée typique de 32 ms, avec un recouvrement inter-trames de 50 % typiquement. On

ouvre pour cela une fenêtre d'analyse représentée par les blocs 402 et 403.

L'opération d'égalisation adaptée est mise en œuvre par un filtre RIF 300 dont les coefficients sont calculés à chaque trame d'activité vocale par la chaîne de traitement illustrée sur les figures 16 et 17.

Le calcul de ces coefficients correspond au calcul de la réponse impulsionnelle du filtre à partir du module de la réponse fréquentielle.

Le spectre à long terme de $x(n)$, γ_x , est d'abord calculé (à partir de l'instant initial de fonctionnement) sur une fenêtre temporelle croissant de 0 à une durée T d'activité vocale (typiquement 4 s), puis ajusté récursivement à chaque trame d'activité vocale, ce qui se traduit par la formule générique suivante :

$$\gamma_x(f, n) = \alpha(n) |X(f, n)|^2 + (1 - \alpha(n)) \gamma_x(f, n-1), \quad (0.10)$$

où $\gamma_x(f, n)$ est le spectre à long terme de x à la n ème trame d'activité vocale, $X(f, n)$ la transformée de Fourier de la n ème trame d'activité vocale, et $\alpha(n)$ est défini par la relation (0.11). En notant N le nombre de trames dans la durée T ,

$$\alpha(n) = \frac{1}{\min(n, N)}. \quad (0.11)$$

Ce calcul est opéré par les blocs 421, 422, 423.

On calcule ensuite, à partir de ce spectre à long terme le cepstre partiel C_p , selon la relation (0.4) mise en œuvre par les blocs de traitement 424, 425, 426.

Le pitch moyen $\overline{F_0}$ est estimé par le bloc de traitement 412 à chaque trame voisée selon la formule :

$$\overline{F_0}(m) = \alpha(m) F_0(m) + (1 - \alpha(m)) \overline{F_0}(m-1), \quad (0.12)$$

où $F_0(m)$ est le pitch de la mième trame voisée et est calculé par le bloc 411 selon une méthode appropriée de l'état de l'art (par exemple méthode de l'autocorrélation, avec détermination du voisement par comparaison de l'autocorrélation normalisée à un seuil [UIT-T/G.729, 1996]).

Ainsi, à chaque trame d'activité vocale, on dispose d'un nouveau vecteur x de composantes le pitch moyen et les coefficients 1 à L du cepstre partiel, auquel on applique la fonction discriminante a définie à partir du corpus d'apprentissage. Ce traitement est mis en œuvre par le bloc 413. Le locuteur est alors affecté à la classe q de score discriminant minimal.

Le module en dB de la réponse fréquentielle de l'égaliseur adapté restreinte à la bande F_1 - F_2 , noté $|EQ|_{dB[F_1-F_2]}$, est calculé selon l'une des deux méthodes suivantes :

La première méthode (figure 16) consiste à calculer $|EQ|_{F_1-F_2}$ selon l'équation (0.3), où $\gamma_{ref}(f)$ est le spectre de référence de la classe du locuteur (transformée de Fourier du centre de la classe). Cette méthode de calcul est mise en œuvre dans cette variante représentée sur la figure 16 avec les opérateurs 414a, 428a, 427 et 429.

La deuxième méthode (figure 17) consiste à transcrire l'équation (0.3) dans le domaine du cepstre partiel, puisque l'on dispose du cepstre partiel de la sortie x du pré-égaliseur, nécessaire au classement du locuteur. Ainsi, l'équation (0.3) devient :

$$C_{eq}^p = C_{ref}^p - C_x^p - C_{s_{rx}}^p - C_{l_{rx}}^p, \quad (0.13)$$

où C_{eq}^p , C_x^p , $C_{s,rx}^p$ et $C_{l,rx}^p$ sont les cepstres partiels respectifs de l'égaliseur adapté, de la sortie x du pré-égaliseur, du système de réception et de la ligne de réception, C_{ref}^p étant le cepstre partiel de référence, centre de la classe du locuteur. Les cepstres partiels sont calculés comme indiqué précédemment, en sélectionnant la bande de fréquences F1-F2. Ce calcul est effectué uniquement pour les coefficients 1 à 20, les coefficients suivants étant inutiles car représentatifs d'une finesse spectrale qui sera éliminée par la suite.

Les 20 coefficients du cepstre partiel de l'égaliseur adapté sont obtenus par les opérateurs 414b et 428b selon la relation (0.13).

Le bloc de traitement 441 complète ces 20 coefficients par des zéros, les symétrise et calcule, à partir du vecteur ainsi formé, le module en dB de la réponse fréquentielle de l'égaliseur adapté restreinte à la bande F1-F2 en mettant en œuvre la relation suivante :

$$EQ_{dB|F_1-F_2} = TFD^{-1}(C_{eq}^p). \quad (0.14)$$

Cette réponse est décimée d'un facteur $\frac{3}{4}$ par l'opérateur 442.

Pour les deux variantes qui viennent d'être décrites, les valeurs de $|EQ|$ hors de la bande F1-F2 sont calculées par extrapolation linéaire de la valeur en dB de $|EQ|_{F_1-F_2}$, notée EQ_{dB} par la suite, par le bloc 430 et de la manière suivante :

Pour chaque indice de fréquence k , l'approximation linéaire de EQ_{dB} s'exprime par :

$$\tilde{EQ}_{dB}(k) = a_1 + a_2 k \quad (0.15)$$

Les coefficients a_1 et a_2 sont choisis de manière à minimiser l'erreur quadratique de l'approximation sur l'intervalle F_1 - F_2 , définie par

$$e = \sum_{k=k_1}^{k_2} \left(EQ_{dB}(k) - \tilde{EQ}_{dB}(k) \right)^2 \quad (0.16)$$

5 Les coefficients a_1 et a_2 sont donc définis par :

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} k_2 - k_1 + 1 & \sum_{k=k_1}^{k_2} k \\ \sum_{k=k_1}^{k_2} k & \sum_{k=k_1}^{k_2} k^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{k=k_1}^{k_2} EQ_{dB}(k) \\ \sum_{k=k_1}^{k_2} k EQ_{dB}(k) \end{pmatrix} \quad (0.17)$$

Les valeurs de $|EQ|$, en dB, hors de la bande F_1 - F_2 , sont alors calculées à partir de la formule (0.15).

10 La caractéristique en fréquence ainsi obtenue doit être lissée. Le filtrage étant réalisé dans le domaine temporel, le moyen permettant ce lissage est de multiplier par une fenêtre étroite la réponse impulsionnelle correspondante.

15 La réponse impulsionnelle est obtenue par une opération IFFT appliquée sur $|EQ|$ réalisée par les blocs 431 et 432 suivie d'une symétrisation réalisée par le bloc de traitement 433, de manière à obtenir un filtre causal à phase linéaire. La réponse impulsionnelle résultante est multipliée, opérateur
20 435, par une fenêtre temporelle 434. La fenêtre utilisée est typiquement une fenêtre de Hamming de longueur 31 centrée sur le pic de la réponse impulsionnelle et est appliquée sur la réponse impulsionnelle au moyen de l'opérateur 435.

REVENDICATIONS

1. Procédé de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant une opération d'égalisation sur une bande de fréquence $[F1-F2]$, adaptée à la distorsion réelle de la chaîne de transmission, cette opération étant réalisée au moyen d'un filtre numérique ayant une réponse fréquentielle fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal voix des locuteurs, caractérisé en ce qu'il comprend :

* Préalablement à l'opération d'égalisation du signal voix d'un locuteur en communication:

- La constitution de classes de locuteurs avec une référence de voix par classe,

* Puis, pour un locuteur en communication donné :

- Le classement de ce locuteur c'est-à-dire son affectation à une classe à partir de critères de classement prédéfinis pour lui faire correspondre une référence de voix qui est la plus proche de la sienne,

- L'égalisation du signal numérisé de la voix du locuteur opérée avec comme spectre de référence, la référence de voix de la classe à laquelle ledit locuteur a été affecté.

2. Procédé de correction des déformations spectrales de la voix selon la revendication 1, caractérisé en ce que :

* La constitution de classes de locuteurs comprend :

- Le choix d'un corpus de N locuteurs enregistrés dans des conditions non dégradées et la détermination de leur spectre de fréquence à long terme,

5 - La classification des locuteurs du corpus selon leur cepstre partiel c'est à dire le cepstre calculé à partir du spectre à long terme restreint à la bande d'égalisation [F1-F2] et en appliquant un critère de classification prédéfini sur ces cepstres pour obtenir K classes,

10 - Le calcul du spectre de référence associé à chaque classe de manière à obtenir une référence de voix correspondant à chacune des classes.

15 3. Procédé de correction des déformations spectrales de la voix selon la revendication 2, caractérisé en ce que le spectre de référence sur la bande de fréquences d'égalisation [F1-F2], associé à chaque classe, est calculé par transformée de Fourier du centre de la classe défini par son cepstre partiel.

20 4. Procédé de correction des déformations spectrales de la voix selon la revendication 1, caractérisé en ce que :

* La classification d'un locuteur comprend :

25 - L'utilisation du pitch moyen du signal voix et du cepstre partiel de ce signal comme paramètres de classement,

- L'application d'une fonction discriminante à ces paramètres pour classer ledit locuteur.

30

5. Procédé de correction des déformations spectrales de la voix selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il comprend en outre une étape de pré-égalisation du

signal numérique par un filtre fixe ayant une réponse fréquentielle dans la bande de fréquence [F1-F2], correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison téléphonique.

5

6. Procédé de correction des déformations spectrales de la voix selon l'une quelconque des revendications précédentes, caractérisé en ce que l'égalisation du signal numérisé de la voix d'un locuteur comprend :

- la détection d'une activité vocale sur la ligne pour déclencher un enchaînement de traitements comprenant le calcul du spectre à long terme, la classification du locuteur, le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] et le calcul des coefficients du filtre numérique différenciés selon la classe du locuteur, à partir de ce module,
- la commande du filtre avec les coefficients obtenus,
- le filtrage du signal sortant du pré-égaliseur par ledit filtre.

20

7. Procédé de correction des déformations spectrales de la voix selon la revendication 6, caractérisé en ce que le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] est réalisé par la mise en œuvre de la relation suivante :

25

30

$$|EQ(f)| = \frac{1}{|S_{RX}(f) \cdot L_{RX}(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \quad (0.3)$$

dans laquelle $\gamma_{ref}(f)$ est le spectre de référence de la classe à laquelle appartient ledit locuteur,

et dans laquelle L_RX est la réponse fréquentielle de la ligne de réception, S_RX la réponse fréquentielle du système de réception et $\gamma x(f)$ le spectre à long terme du signal x d'entrée du filtre.

8. Procédé de correction des déformations spectrales de la voix selon la revendication 6, caractérisé en ce que le calcul du module [EQ] de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] est réalisé par la mise en œuvre de la relation suivante :

$$C_{eq}^p = C_{ref}^p - C_x^p - C_{s_{rx}}^p - C_{l_{rx}}^p, \quad (0.13)$$

dans laquelle C_{eq}^p , C_x^p , $C_{s_{rx}}^p$ et $C_{l_{rx}}^p$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C_{ref}^p étant le cepstre partiel de référence, centre de la classe du locuteur ; le module [EQ] restreint à la bande F1-F2 étant calculé par Transformée de Fourier Discrète de C_{eq}^p .

9. Système de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant des moyens d'égalisation adaptée dans une bande de fréquence [F1-F2] qui comportent un filtre numérique (300) dont la réponse fréquentielle est fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme d'un signal voix, caractérisé en ce que ces moyens comprennent en outre :

- des moyens de traitement du signal (400) pour le calcul des coefficients du filtre numérique munis :

- d'un premier bloc de traitements (400A) du signal pour calculer le module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] selon la relation suivante :

$$|EQ(f)| = \frac{1}{|S_{RX}(f).L_{RX}(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \quad (0.3)$$

dans laquelle $\gamma_{ref}(f)$ est le spectre de référence, qui peut être différent d'un locuteur à un autre et qui correspond à une référence de classe prédéterminée à laquelle appartient ledit locuteur, et dans laquelle L_{RX} est la réponse fréquentielle de la ligne de réception, S_{RX} la réponse fréquentielle du système de réception et $\gamma_x(f)$ le spectre à long terme du signal x d'entrée du filtre;

- d'un deuxième bloc de traitements (400B) pour le calcul de la réponse impulsionnelle à partir du module de réponse fréquentielle ainsi calculé, afin de déterminer les coefficients du filtre différenciés selon la classe du locuteur

10. Système de correction des déformations spectrales de la voix selon la revendication 9, caractérisé en ce que le premier bloc de traitements (400A) comprend des moyens (414b, 428b) de calcul du cepstre partiel du filtre égaliseur selon la relation :

$$C_{eq}^p = C_{ref}^p - C_x^p - C_{s_{rx}}^p - C_{l_{rx}}^p, \quad (0.13)$$

dans laquelle C_{eq}^p , C_x^p , $C_{s_{rx}}^p$ et $C_{l_{rx}}^p$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C_{ref}^p étant le cepstre

partiel de référence, centre de la classe du locuteur ;
le module [EQ] restreint à la bande F1-F2 étant calculé
par Transformée de Fourier Discrète de $C^{P_{eq}}$.

5 11. Système de correction des déformations
spectrales de la voix selon la revendication 9 ou 10,
caractérisé en ce que le premier bloc de traitements
comprend un sous-ensemble (420) pour calculer les
coefficients du cepstre partiel d'un locuteur en
10 communication et un deuxième sous-ensemble (410) pour
opérer le classement de ce locuteur, ce deuxième sous
ensemble comprenant un bloc (411) de calcul du pitch F_0 ,
un bloc (412) d'estimation du pitch moyen à partir du
pitch calculé F_0 , et un bloc (413) de classement
15 appliquant une fonction discriminante sur le vecteur x
ayant pour composantes le pitch moyen et les
coefficients du cepstre partiel pour classer ledit
locuteur.

20 12. Système de correction des déformations
spectrales de la voix selon l'une quelconque des
revendications 9 à 11, caractérisé en ce qu'il comprend
un pré-égaliseur (200) et en ce que le signal égalisé à
partir de spectres de référence différenciés selon la
25 classe du locuteur est le signal x de sortie du pré-
égaliseur.

117

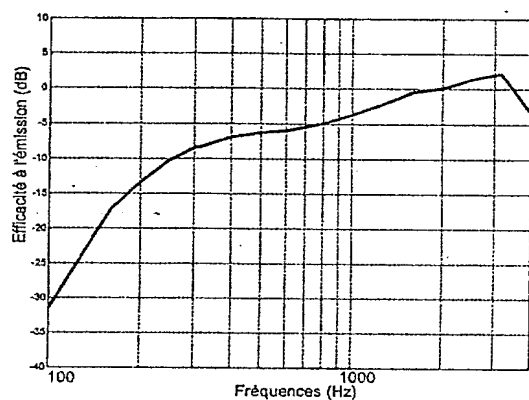
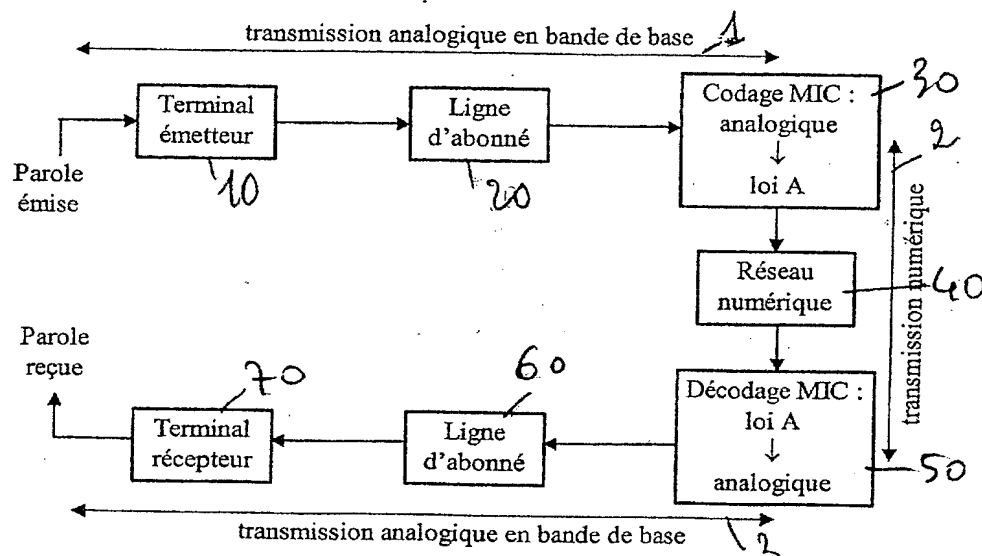


Figure 2

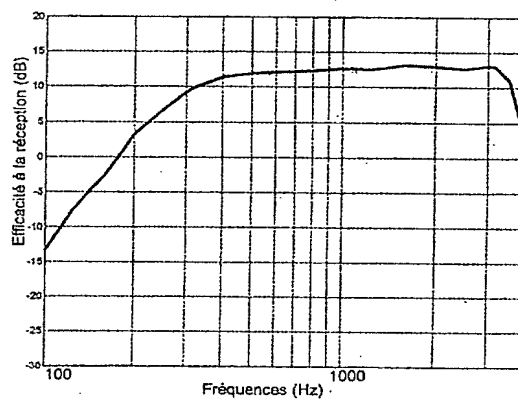


Figure 3

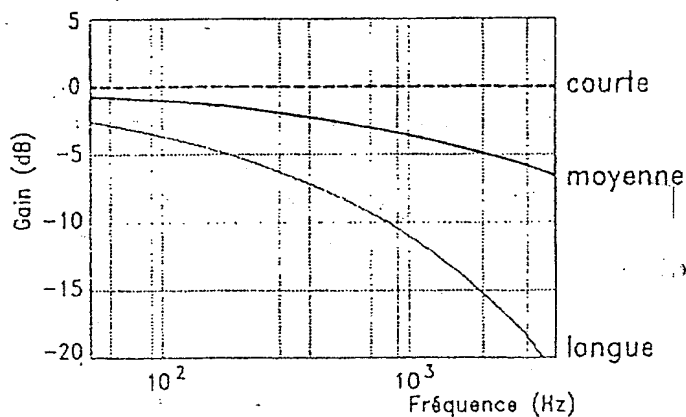


Fig 4

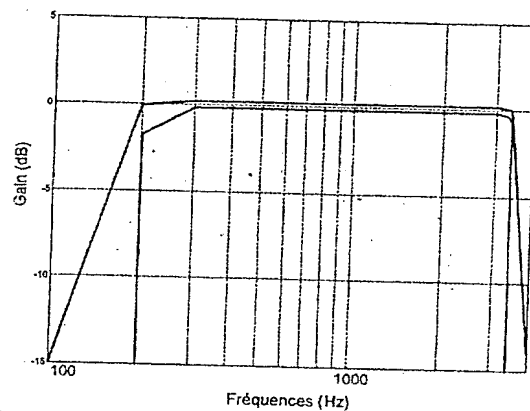


Fig 5

1/7

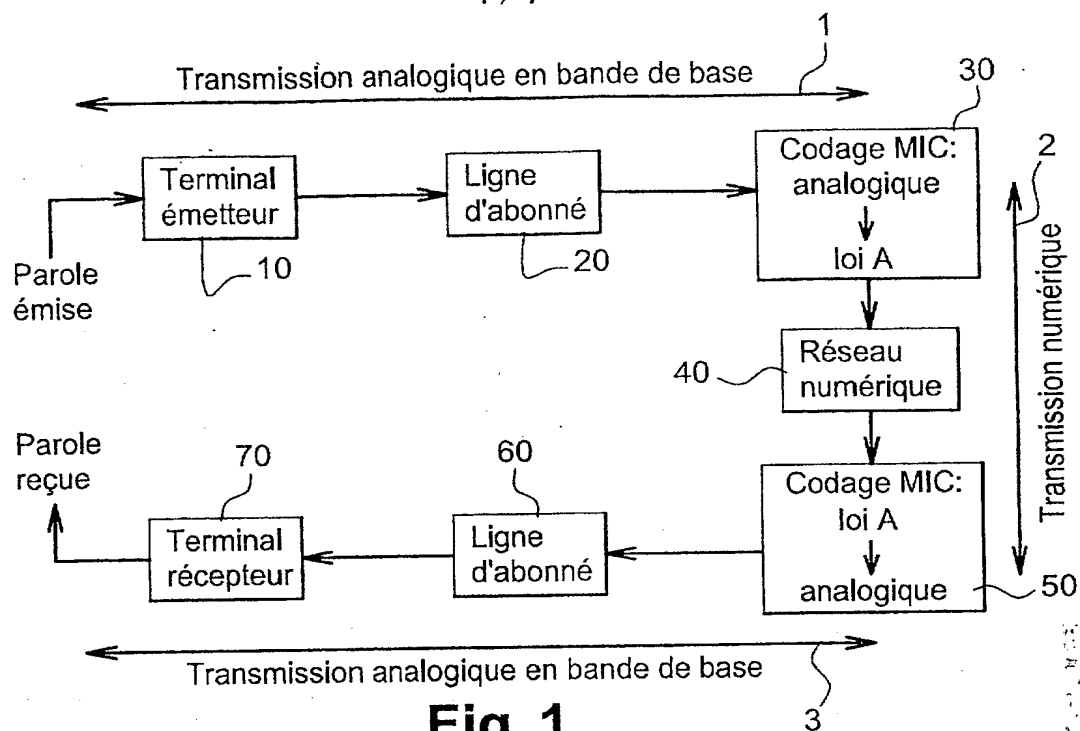


Fig. 1

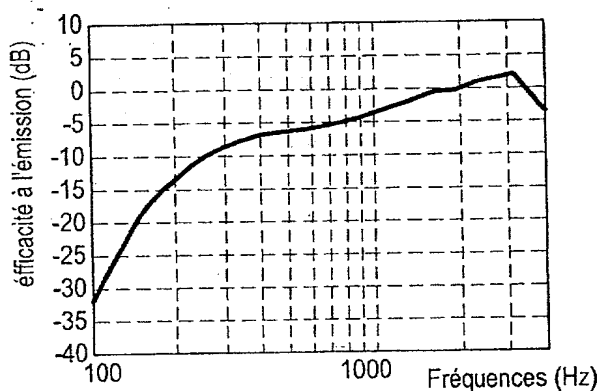


Fig. 2

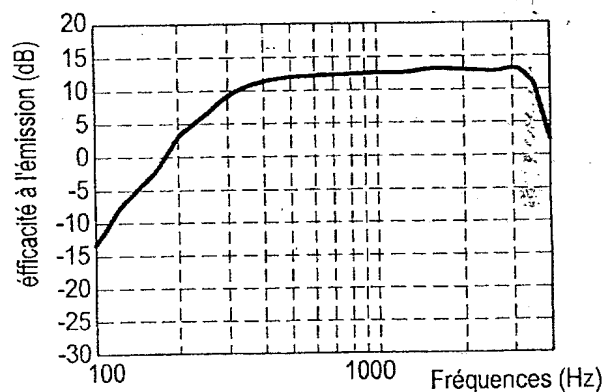


Fig. 3

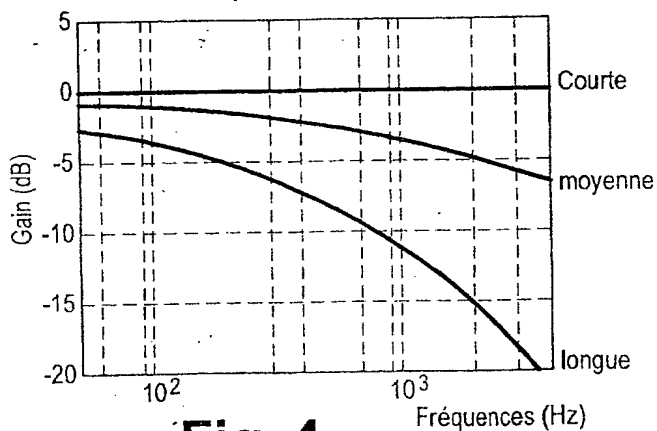


Fig. 4

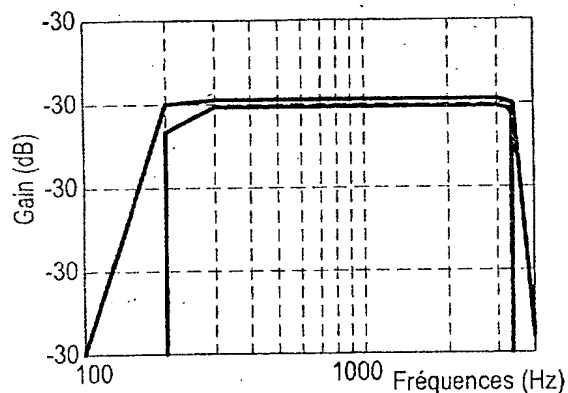


Fig. 5

27

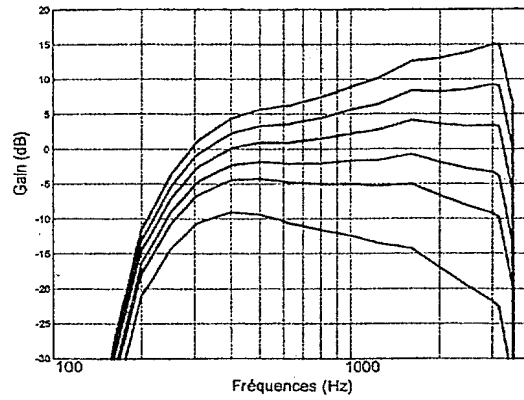


Fig 6.

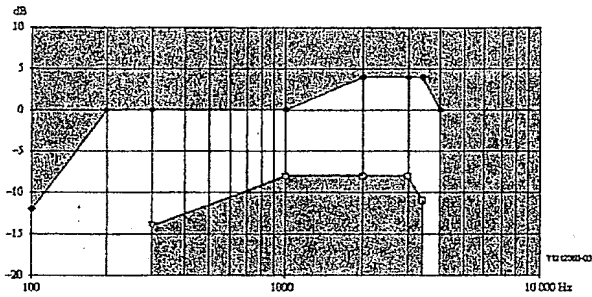


Figure 7

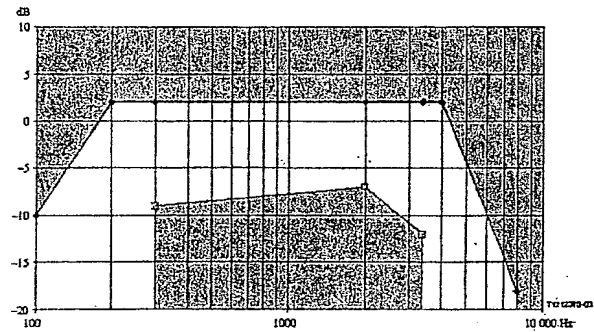


Figure 8

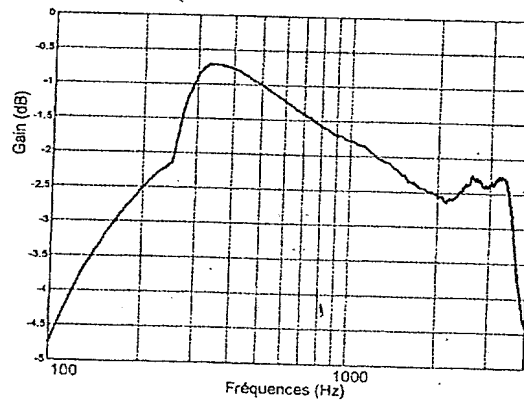


Figure 9

2 / 7

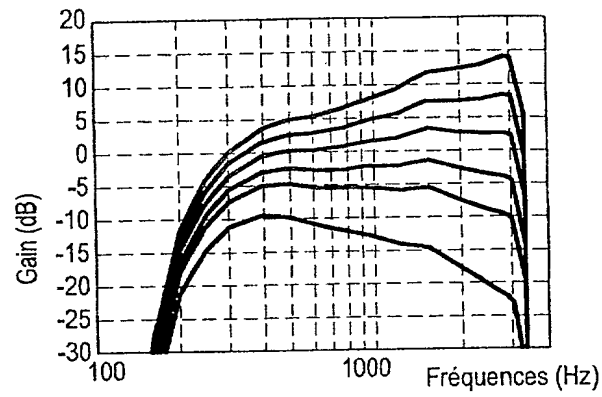


Fig. 6

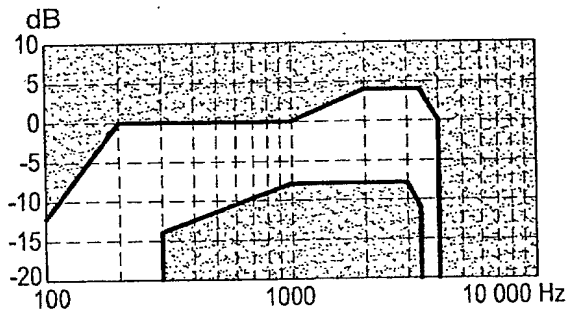


Fig. 7

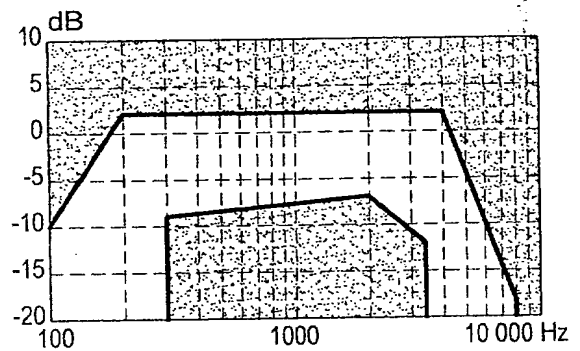


Fig. 8

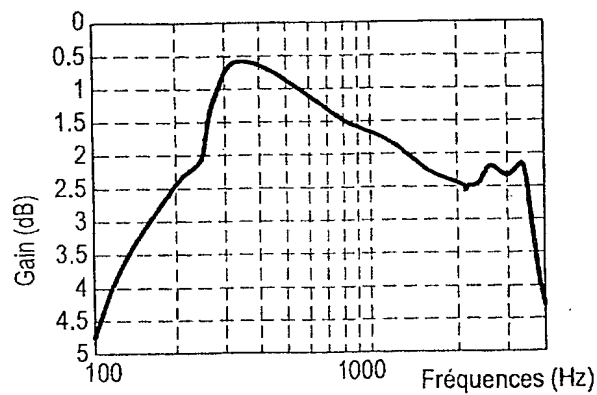


Fig. 9

317

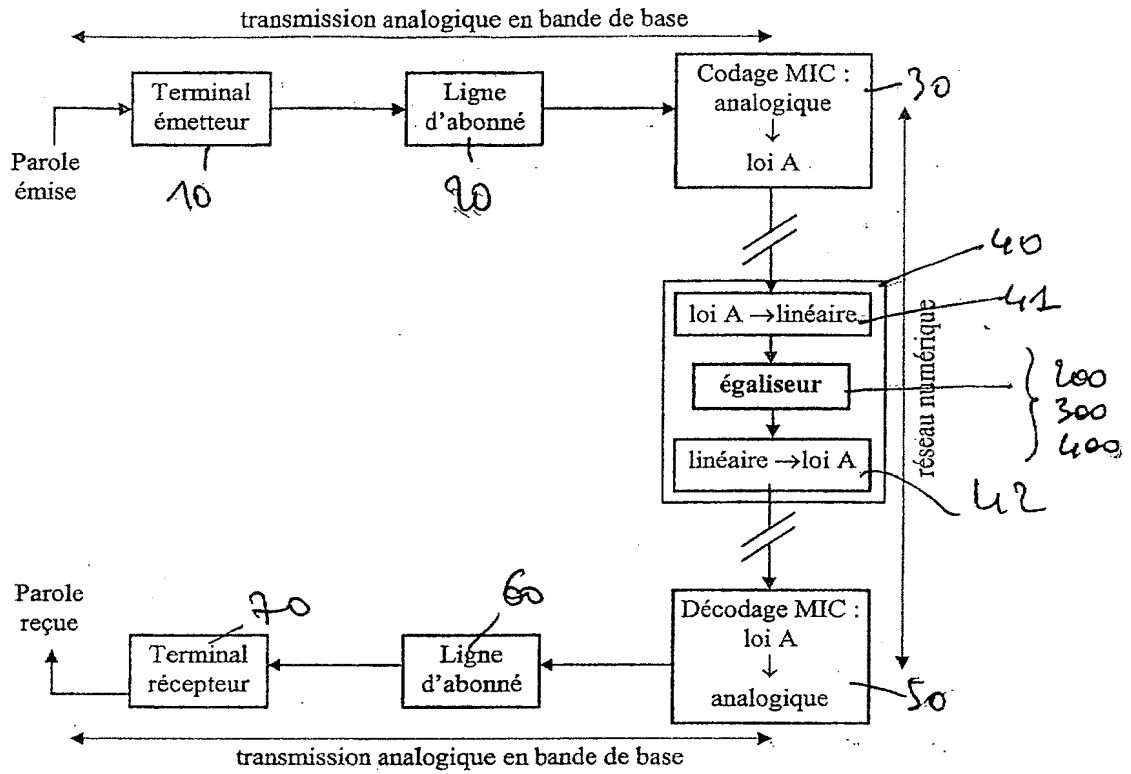


Figure 10

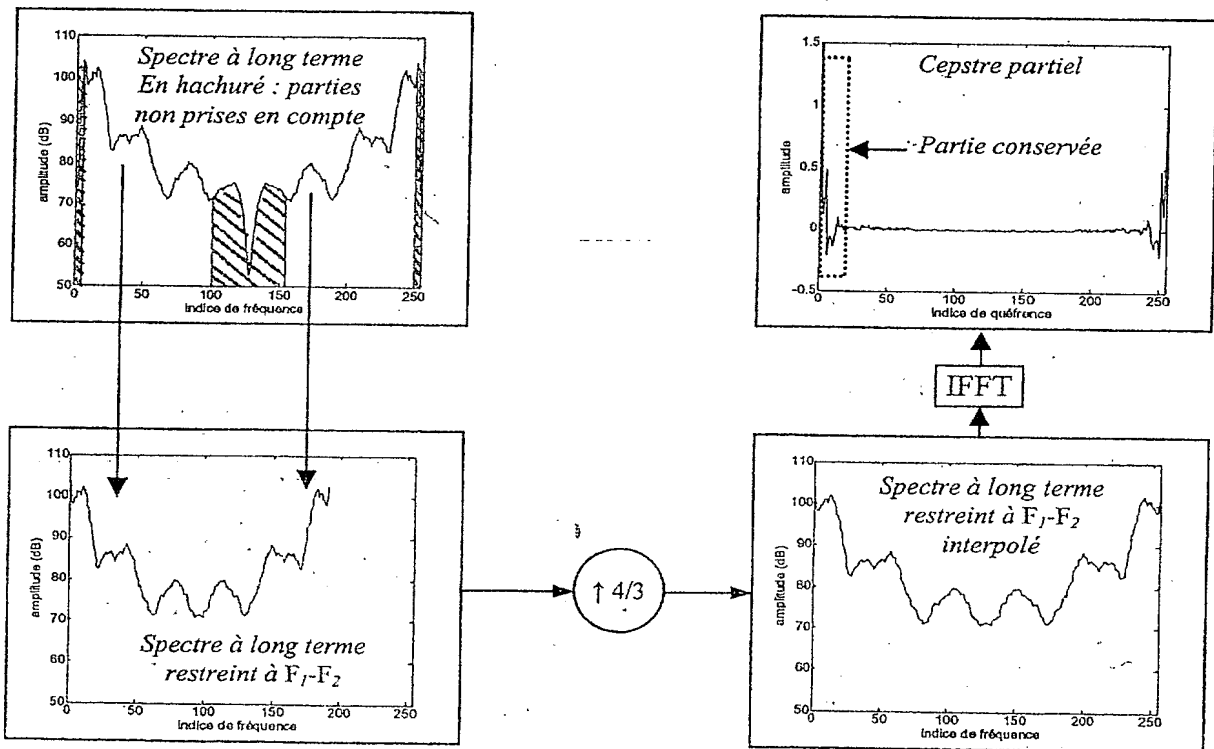
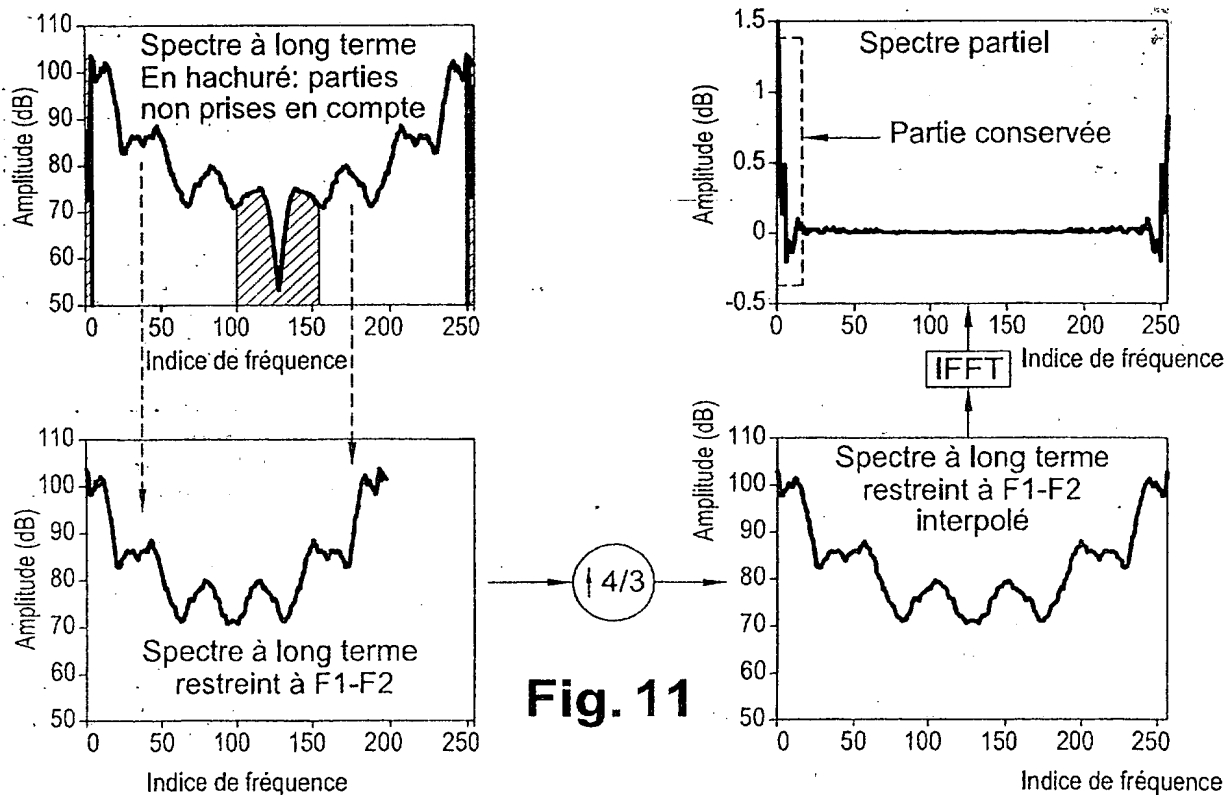
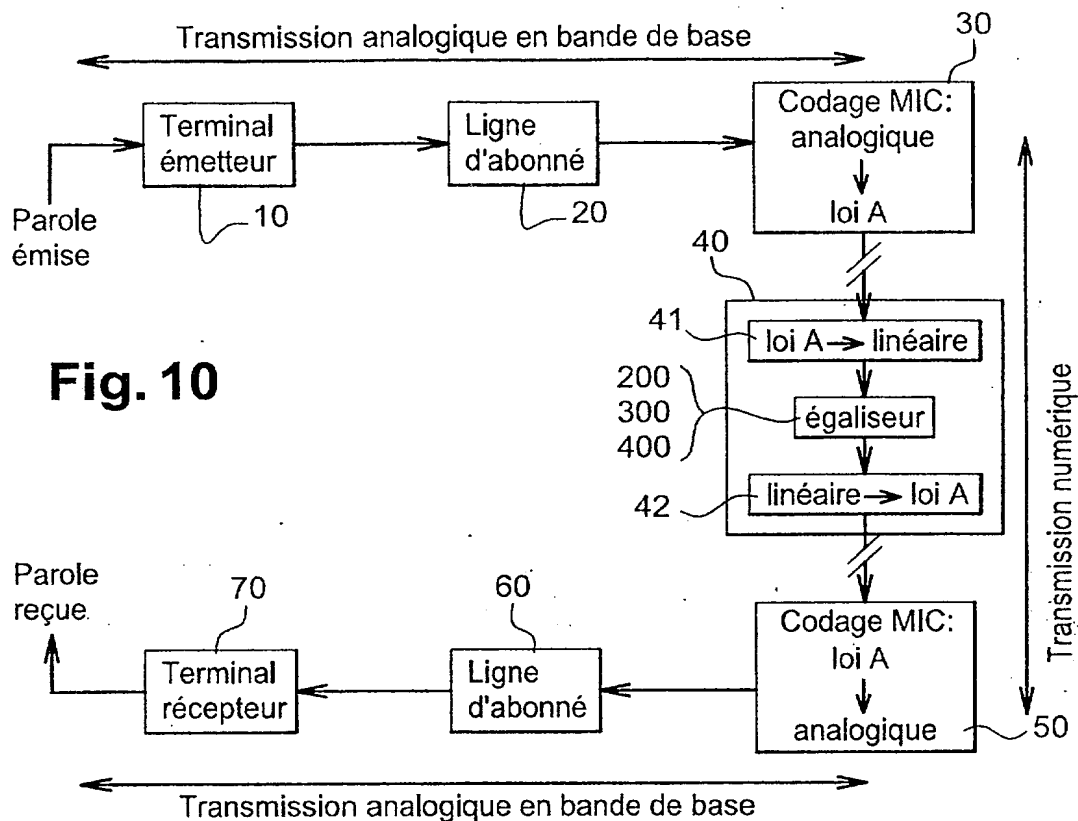


Figure 11

3 / 7



417

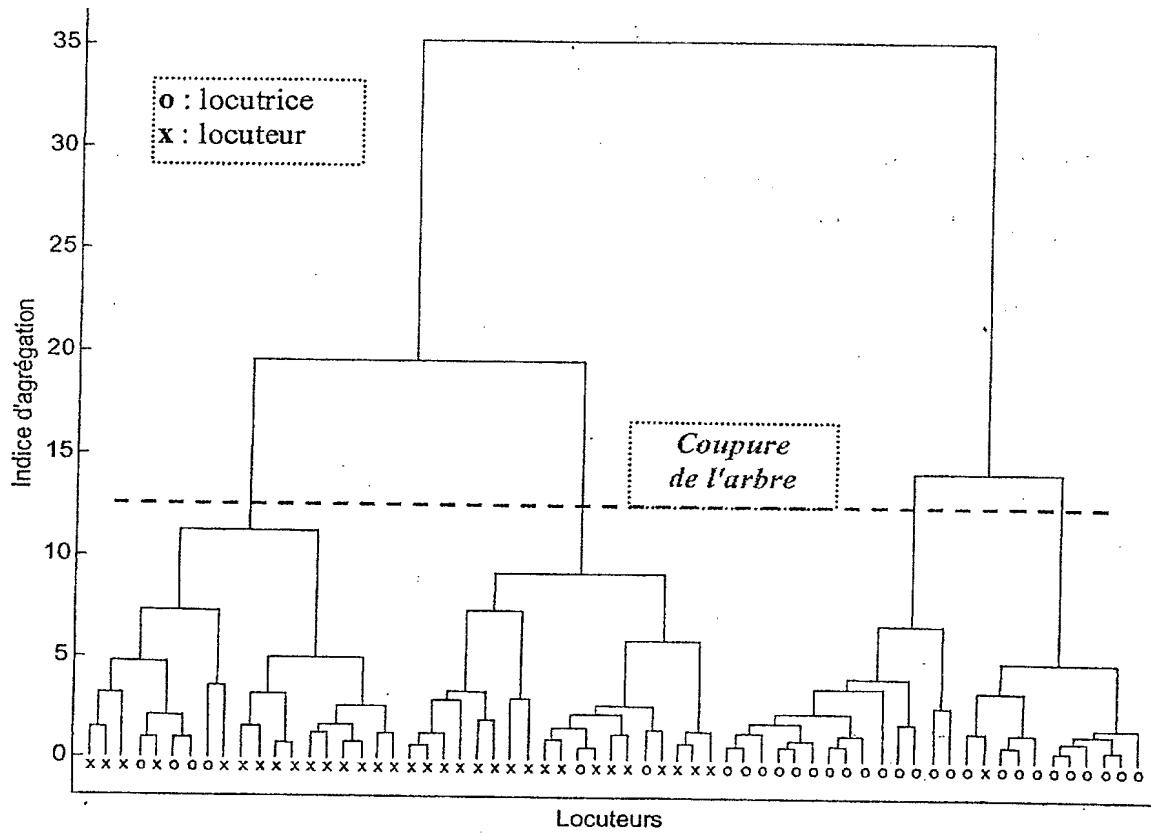


Figure 12 :

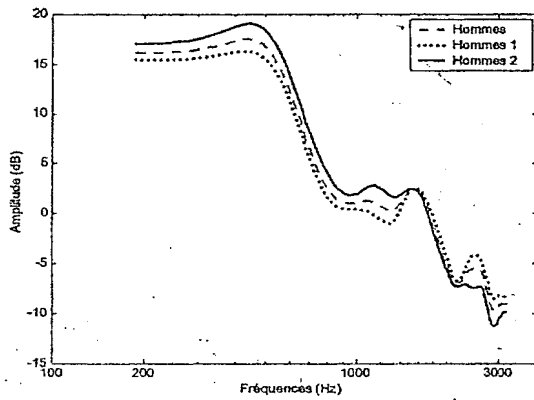


Figure 13a

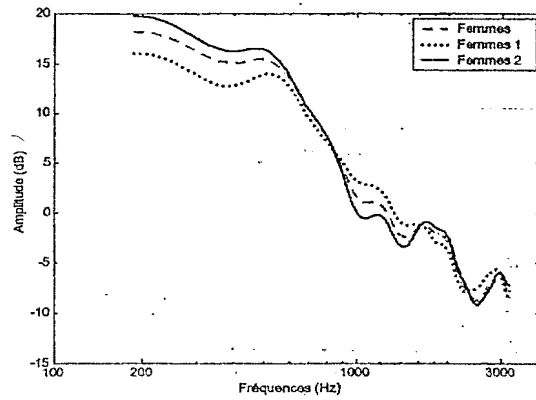


Fig 13b

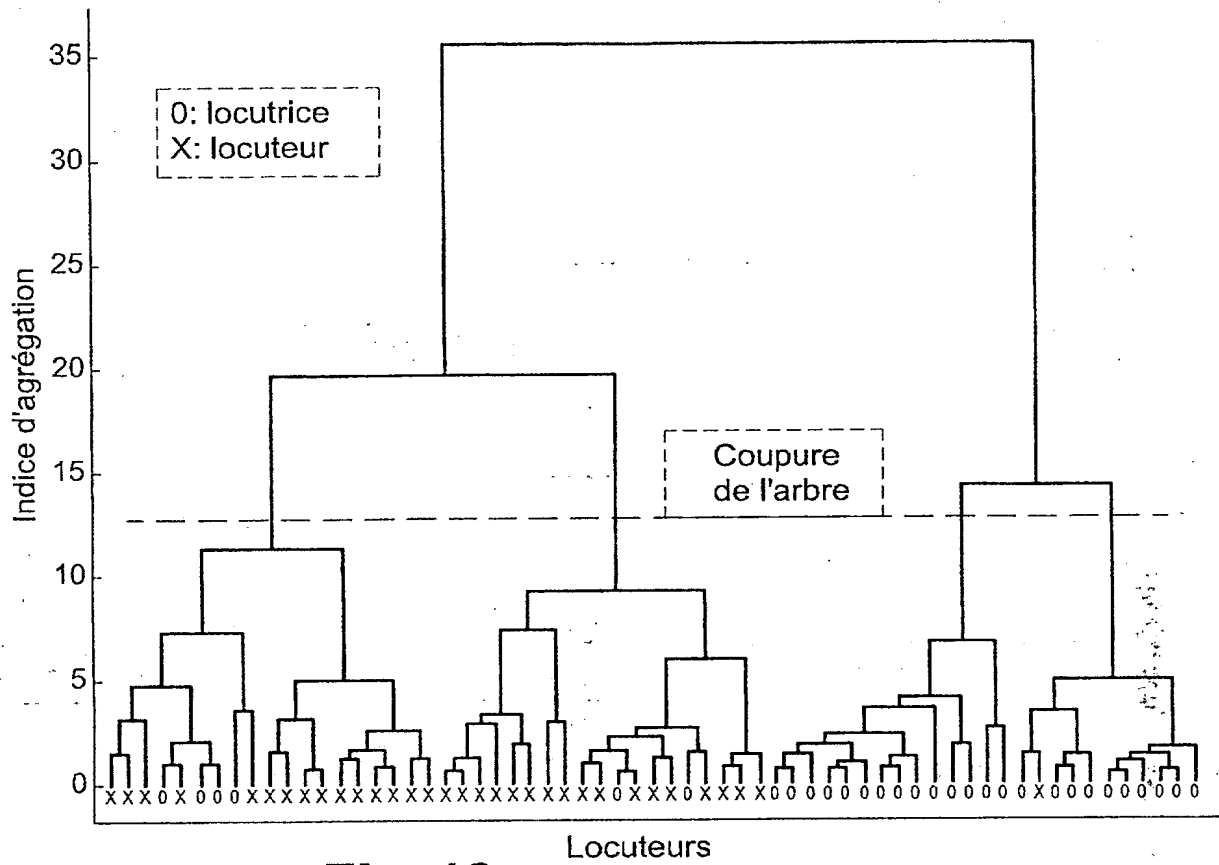


Fig. 12

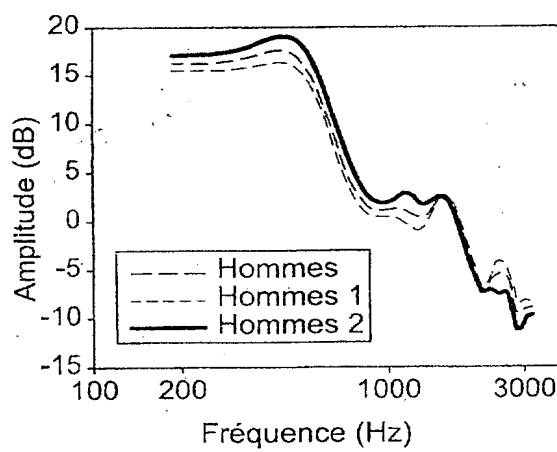


Fig. 13a

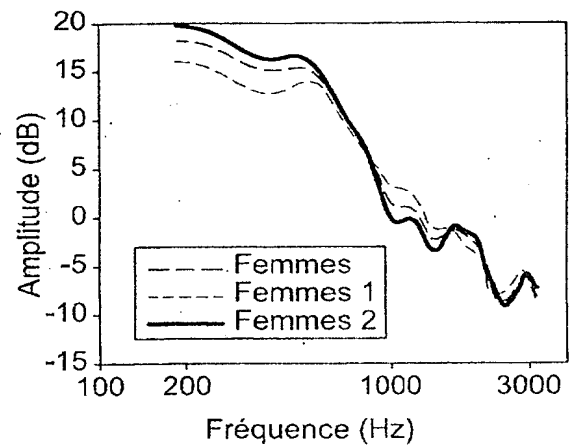


Fig. 13b

5/7

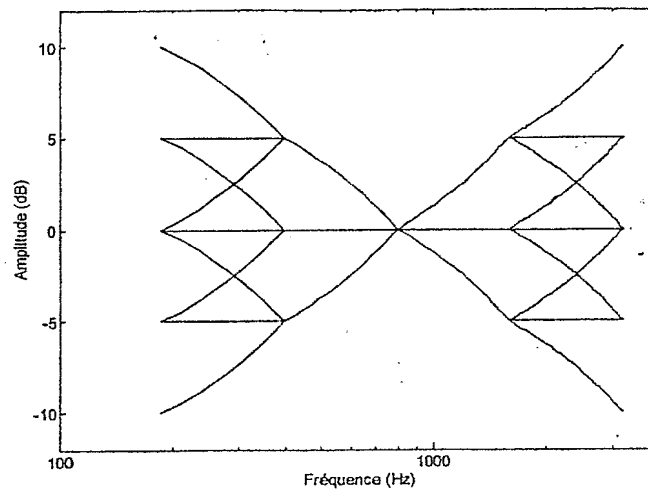


Figure 14

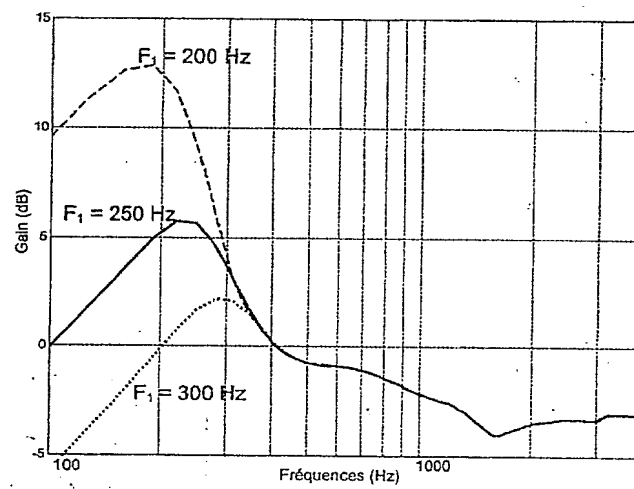


Figure 15

5 / 7

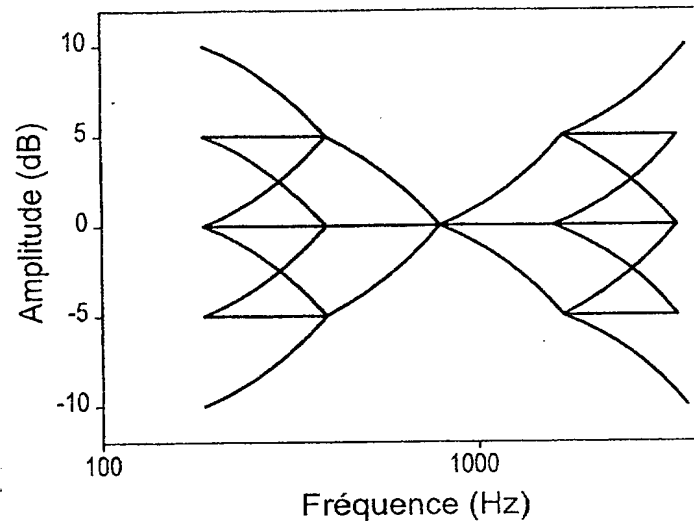


Fig. 14

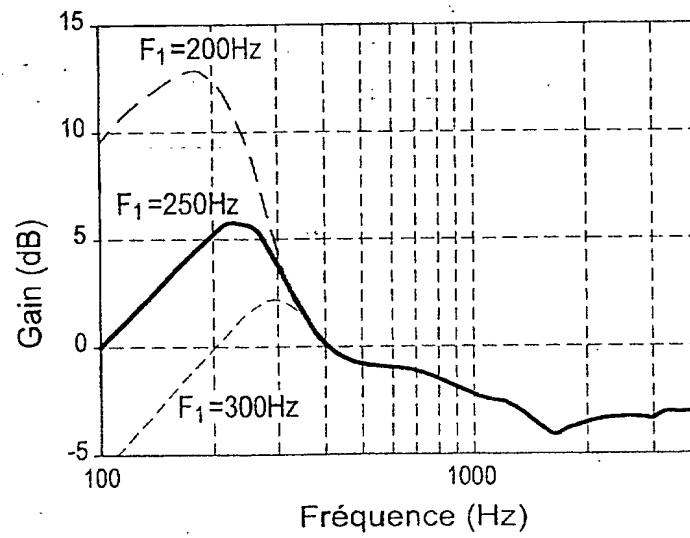
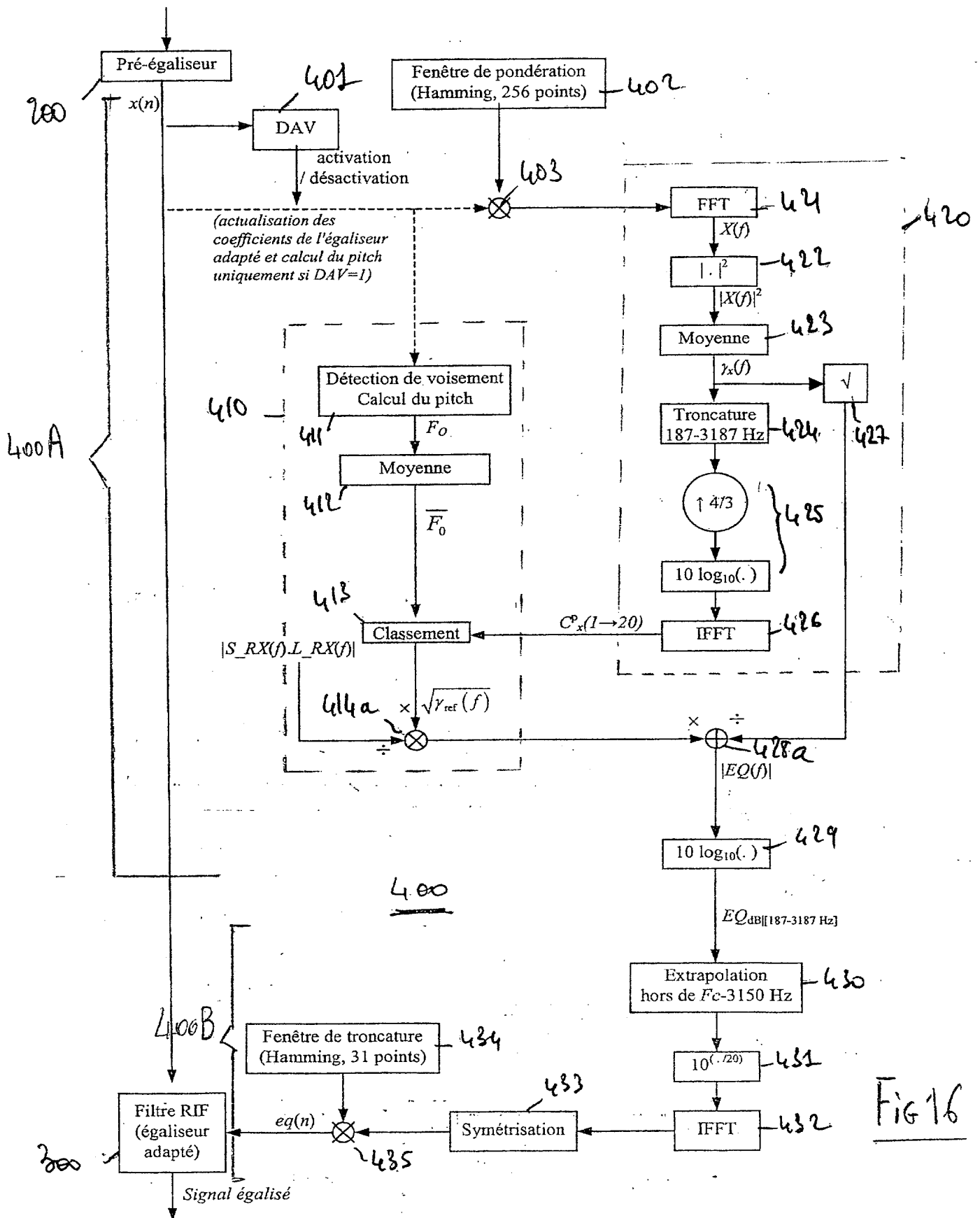


Fig. 15

617



6 / 7

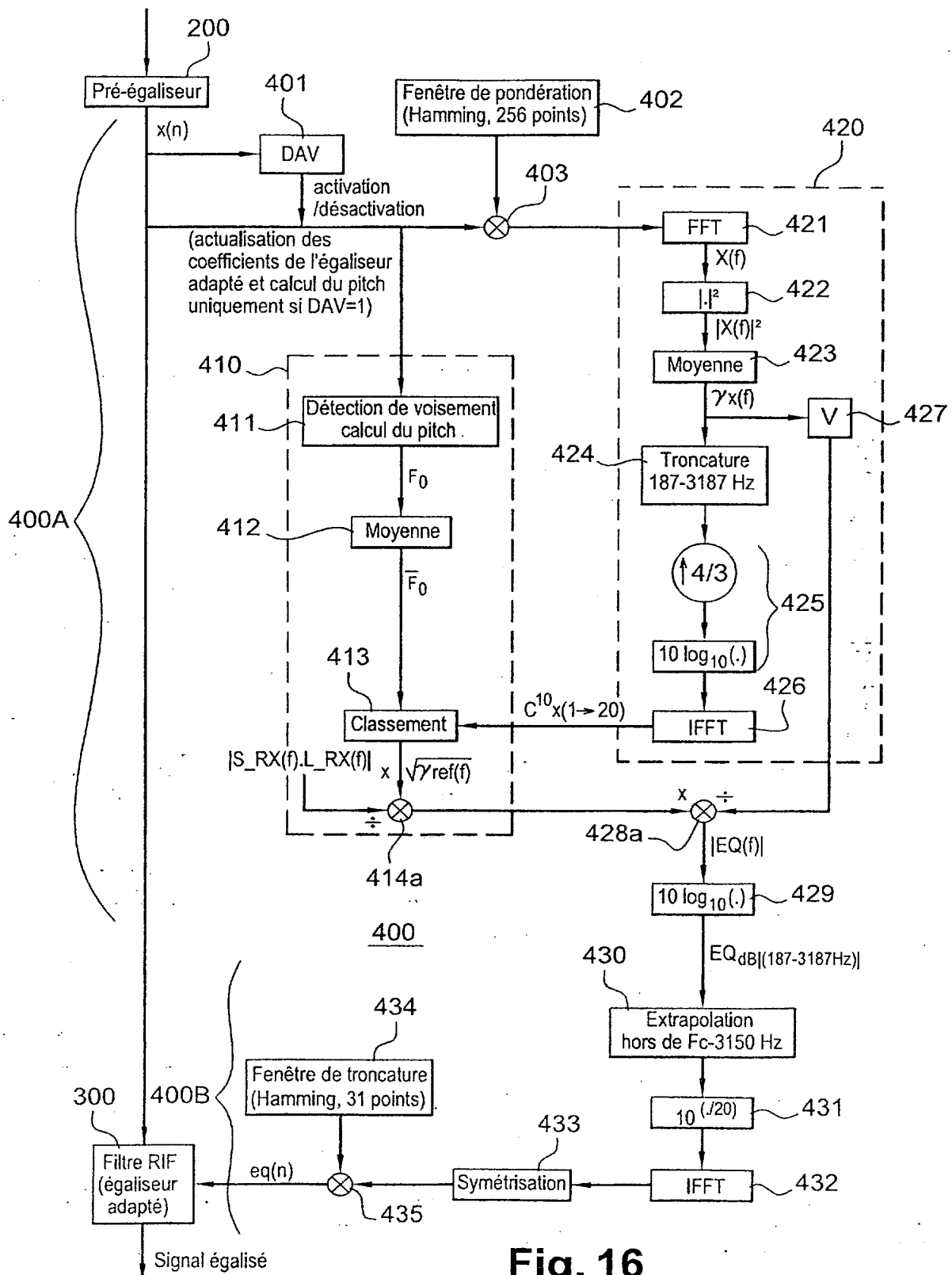
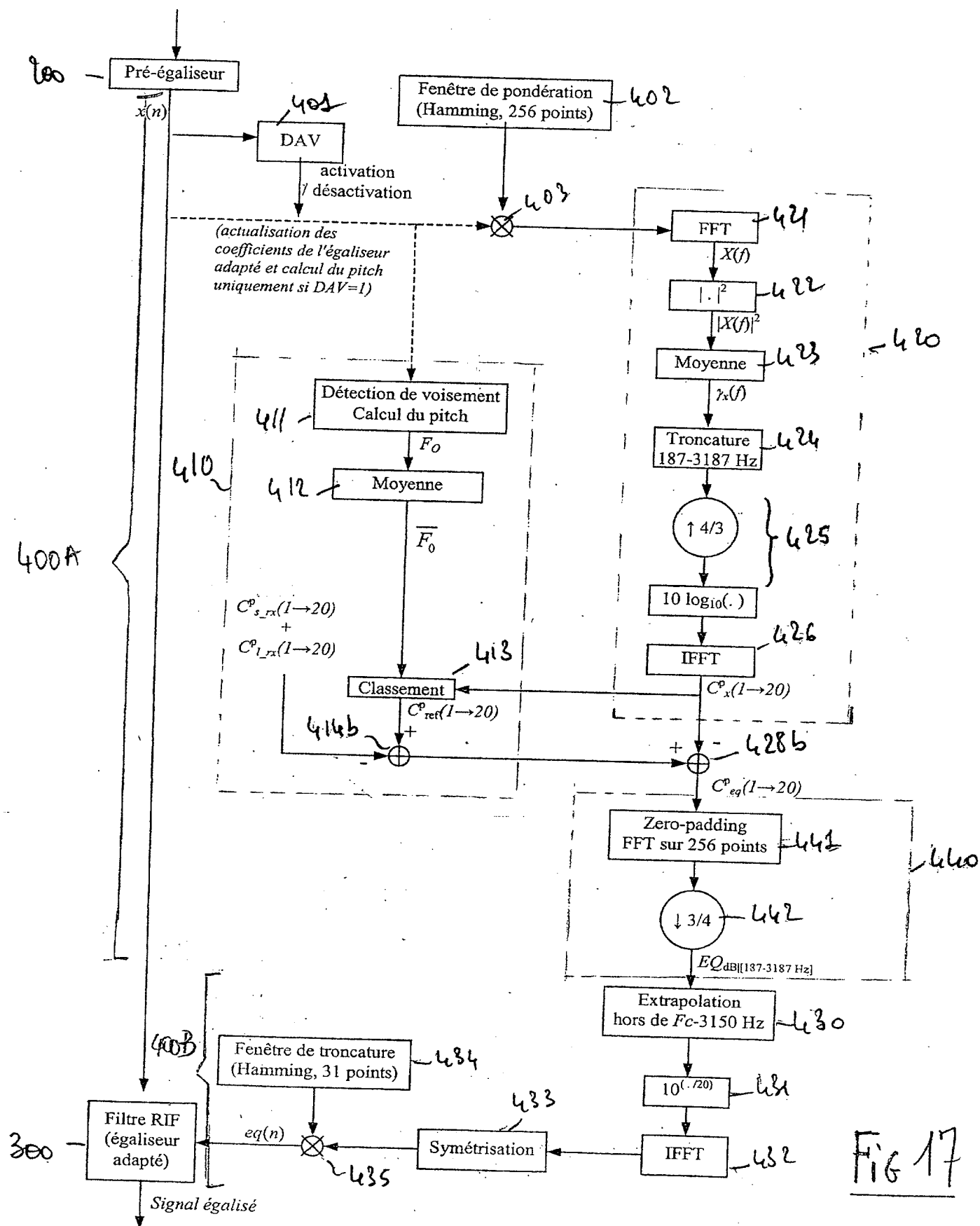


Fig. 16

717



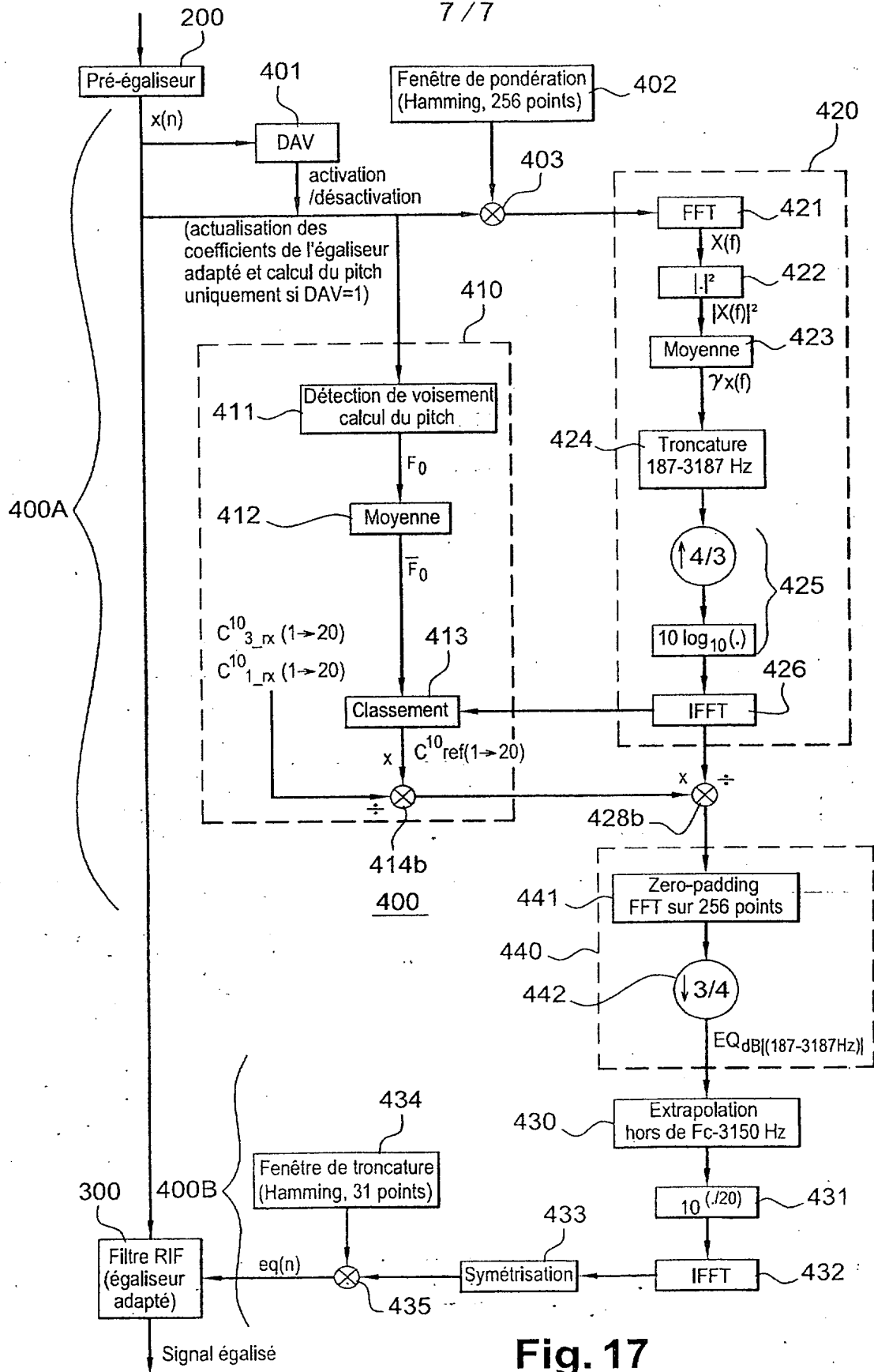


Fig. 17



DÉPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

BREVET D'INVENTION**CERTIFICAT D'UTILITÉ**

Code de la propriété intellectuelle - Livre VI



N° 11235*02

DÉSIGNATION D'INVENTEUR(S) Page N° 1.. / 1..

(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

08 113 W / 260899

Vos références pour ce dossier (facultatif)		0079-FT04453	
N° D'ENREGISTREMENT NATIONAL		02 15618	
TITRE DE L'INVENTION (200 caractères ou espaces maximum) PROCÉDE ET SYSTÈME DE CORRECTION MULTI-REFERENCES DES DEFORMATIONS SPECTRALES DE LA VOIX INTRODUITES PAR UN RESEAU DE COMMUNICATION.			
LE(S) DEMANDEUR(S) : FRANCE TELECOM 6, Place d'Alleray 75015 PARIS France			
DESIGNE(NT) EN TANT QU'INVENTEUR(S) : (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).			
Nom		MAHE	
Prénoms		Gaël	
Adresse	Rue	Résidence Beauchamp 74, rue de Tréguier - Bâtiment H	
	Code postal et ville	22300	LANNION
Société d'appartenance (facultatif)			
Nom		GILLOIRE	
Prénoms		André	
Adresse	Rue	7, Rue Yann Péron	
	Code postal et ville	22300	LANNION
Société d'appartenance (facultatif)			
Nom			
Prénoms			
Adresse	Rue		
	Code postal et ville		
Société d'appartenance (facultatif)			
DATE ET SIGNATURE(S) DU (DES) DEMANDEUR(S) OU DU MANDATAIRE (Nom et qualité du signataire) BORIN Lydie Mandataire N° 94-0506 A. P. I. Conseil Le 03 Décembre 2002			

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire.
Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.